

# Maximal compression of the redshift-space galaxy power spectrum and bispectrum

Davide Gualdi,<sup>1</sup>★ Marc Manera,<sup>1,2,3</sup> Benjamin Joachimi<sup>1</sup> and Ofer Lahav<sup>1</sup>

<sup>1</sup>Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>2</sup>Centre for Mathematical Sciences, DAMTP, Cambridge University, Wilberforce Rd, Cambridge CB3 0WA, UK

<sup>3</sup>Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain

Accepted 2018 January 28. Received 2018 January 25; in original form 2017 September 11

## ABSTRACT

We explore two methods of compressing the redshift-space galaxy power spectrum and bispectrum with respect to a chosen set of cosmological parameters. Both methods involve reducing the dimension of the original data vector (e.g. 1000 elements) to the number of cosmological parameters considered (e.g. seven) using the Karhunen–Loève algorithm. In the first case, we run MCMC sampling on the compressed data vector in order to recover the 1D and 2D posterior distributions. The second option, approximately 2000 times faster, works by orthogonalizing the parameter space through diagonalization of the Fisher information matrix before the compression, obtaining the posterior distributions without the need of MCMC sampling. Using these methods for future spectroscopic redshift surveys like DESI, *Euclid*, and PFS would drastically reduce the number of simulations needed to compute accurate covariance matrices with minimal loss of constraining power. We consider a redshift bin of a DESI-like experiment. Using the power spectrum combined with the bispectrum as a data vector, both compression methods on average recover the 68 per cent credible regions to within 0.7 per cent and 2 per cent of those resulting from standard MCMC sampling, respectively. These confidence intervals are also smaller than the ones obtained using only the power spectrum by 81 per cent, 80 per cent, and 82 per cent respectively, for the bias parameter  $b_1$ , the growth rate  $f$ , and the scalar amplitude parameter  $A_s$ .

**Key words:** methods: analytical – methods: data analysis – methods: statistical – cosmological parameters – large-scale structure of Universe – cosmology: miscellaneous.

## 1 INTRODUCTION

In recent years, the number of available large data sets to be used for cosmological analysis has drastically increased (*Planck*,<sup>1</sup> Ade et al. 2014; Sloan Digital Sky Survey,<sup>2</sup> Eisenstein et al. 2011; DES, The Dark Energy Survey Collaboration 2005<sup>3</sup>) and will do even more so in the near future (DESI,<sup>4</sup> Levi et al. 2013; *Euclid*,<sup>5</sup> Laureijs et al. 2011; PFS,<sup>6</sup> Takada et al. 2014). Improving current analysis techniques to extract as much information as possible from these catalogues has become highly relevant. Up to now, most work has been done using two points (2pt) statistics like the two points correlation function or its Fourier transform, the power spectrum.

However, gravity increases the level of non-linearity in the matter distribution field, creating non-Gaussian features which are not constrained by the sole use of 2pt statistics. Higher order statistics like the three points (3pt) correlation function or its Fourier transform, the bispectrum, have already been studied in the past in order to capture the non-Gaussian part of the statistical information contained in the large-scale structure galaxy field. The first measurements of the 3pt correlation function and the bispectrum on a galaxy catalogue were performed by Peebles & Groth (1975), Groth & Peebles (1977), and Fry & Seldner (1982). Fry (1984) studied the relation between the cosmological and bias parameters, modelling the relation between luminous and dark matter, which affects the amplitude and the shape of the bispectrum. Integrated 3pt statistics like the skewness were introduced in order to reduce the complexity and number of modelling parameters (Peebles 1980; Fry & Scherrer 1994; Bernardeau 1994; Juszkiewicz, Hellwing & van de Weygaert 2013). The modelling of redshift-space distortions into the 3pt statistics was later introduced and studied by Matarrese, Verde & Heavens (1997a), Verde et al. (1998), Heavens, Matarrese & Verde (1998), Scoccimarro et al. (1998), and Scoccimarro (2000). Different 3pt statistics have also been proposed as useful tools to

\* E-mail: [davide.gualdi.14@ucl.ac.uk](mailto:davide.gualdi.14@ucl.ac.uk)

<sup>1</sup> <http://sci.esa.int/planck/>

<sup>2</sup> <http://www.sdss3.org/surveys/boss.php>

<sup>3</sup> <https://www.darkenergysurvey.org>

<sup>4</sup> <http://desi.lbl.gov>

<sup>5</sup> <http://sci.esa.int/euclid/>

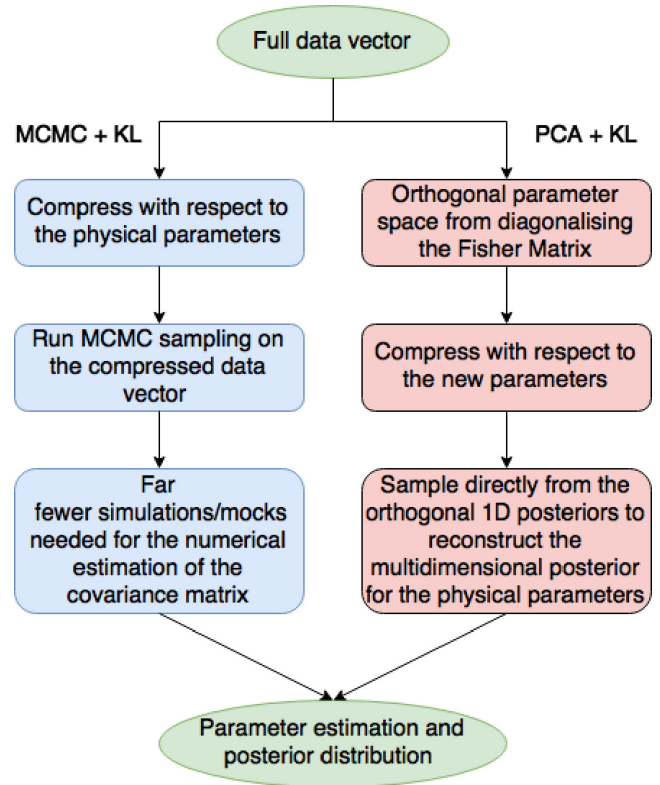
<sup>6</sup> <http://pfs.ipmu.jp>

quantify deviations from GR (Borisov & Jain 2009; Bernardeau & Brax 2011) and to measure primordial non-Gaussianities (Fry & Scherrer 1994; Gangui et al. 1994; Verde et al. 2000; Liguori et al. 2010; Tellarini et al. 2016). In the last 20 yr, these statistics have received increasing attention especially for the benefit of lifting degeneracies present in 2pt statistics between cosmological and nuisance parameters in data sets produced by spectroscopic redshift surveys like BOSS. The most recent measurement on this data set for the bispectrum was made by Gil-Marín et al. (2016) and for the 3pt correlation function by Slepian et al. (2015), obtaining also the first significant BAO detection using 3pt statistics (Slepian et al. 2017).

When studying the power spectrum and the bispectrum, it is necessary to subdivide the range of possible  $\mathbf{k}$ -vectors in Fourier space into bins. By doing this, one defines the number of possible elements for both the power spectrum and bispectrum data vectors. While in Fourier space, the number of possible elements for the power spectrum data vector is the same as the number of bins, for the bispectrum, it corresponds to the number of triangles that can be formed by triplets of the available  $\mathbf{k}$ -vectors (less than  $\sim n_{\text{bins}}^3$ ). This difference becomes even larger when redshift-space distortions (Kaiser 1987), described in Section 2.2, are included in the analysis. In addition, as explained in Appendix B, not only the modulus of the three  $\mathbf{k}$ -vectors but also the orientation of the triangle with respect to the line of sight becomes relevant. Consequently, a fine binning in  $\mathbf{k}$ -space corresponds to a very large number of possible triangles. While for the power spectrum in redshift space, there are  $\sim n_{\text{bins}}^2$  possible data-vector elements, for the bispectrum there are slightly less than  $\sim n_{\text{bins}}^5$  possible triplets, since the vectors need to satisfy the triangle condition.

The problem is that when doing parameter estimation using a data vector for a given statistic, the corresponding covariance matrix is needed. This can either be computed analytically, which requires the evaluation of several multidimensional integral expressions, or it can be estimated from simulations. However, in order to obtain a precise and accurate estimate of the covariance using numerical simulations, the number of realizations must be larger than the number of elements of the data vector (Hartlap, Simon & Schneider 2007; Taylor & Joachimi 2014). Therefore, numerically estimating the covariance matrix of a 3pt statistic from simulations could become very expensive in terms of simulations required to predict it accurately. Indeed, one usually sacrifices the stronger constraining power achievable by considering a larger number of triangle configurations for a more accurate estimate of the covariance matrix obtained using only a subset of all the possible triangles. Compressing the original data vector is an efficient method to avoid losing access to part of the constraining power contained in the 3pt statistics. Alternatives to the full bispectrum have been proposed in the last few years. For example Schmittfull, Baldauf & Seljak (2015) substituted to the tree level matter bispectrum, functions of the cross-power spectrum between  $\delta^2(\mathbf{x})$  and  $\delta$ . The performances of some of them, including the modal decomposition proposed by Schmittfull, Regan & Shellard (2013), have been recently studied by Byun et al. (2017). We will compare their results with ours in the conclusions.

Therefore, one can reduce the original number of data-vector elements either by performing a selection of triangles based on some criteria or by compressing the data vector. This paper is about this second option, where we present an application of a linear compression mechanism by Karhunen and Loève (K-L) used in Tegmark, Taylor & Heavens (1997) for the first time in a cosmological context to compress with respect to multiple parameters. The



**Figure 1.** Diagram highlighting the two compression methods presented in this work, including the respective advantages with respect to the uncompressed data vector.

underlying principle is to assign a vector, such that the weights are proportional to the sensitivity of each element to the variation of a model parameter. When the covariance matrix can be assumed to be parameter independent, the dimension of the compressed data vector corresponds to the number of considered model parameters.

We use this prescription in two ways to do parameter inference, which are summarized in Fig. 1. In our first method, we run an MCMC sampling for the compressed data vector (hereafter MCMC + KL). This has the immediate and appealing benefit for a cosmological survey of requiring fewer simulations to estimate precisely the covariance matrix in the case of a long data vector like the bispectrum (with  $\gtrsim 10^3$  triangles) to just 1 data-vector element for each cosmological parameter. Indeed, in order to have a reliable estimate of the covariance matrix, one needs a significantly larger number of mocks than of data-vector elements. For example, in the DR12 BOSS paper studying the bispectrum (Gil-Marín et al. 2016), the number of triangles used was limited to 825 out of the 6391 possible ones, since only 2048 mocks were available (Kitaura et al. 2016).

In our second method, before compressing, we orthogonalize the parameter space by diagonalizing the Fisher information matrix (principal component analysis, hereafter PCA + KL). This enables sampling from 1D posterior distributions of the new set of parameters, to recover the full multidimensional posterior distribution for the original physical set of parameters without the need of an MCMC sampling. This proves to be  $\sim 10^3$  times faster than an MCMC run, in particular because the linear matter power spectrum needs to be recomputed far fewer times.

In addition to being competitive with the standard likelihood approach, this method could be also very useful for future cosmological surveys in order to test and forecast the constraining power

of the expected data set for several different combinations of cosmological and nuisance (e.g. bias) parameters. We apply our analysis to the redshift-space galaxy bispectrum and to the joint data vector formed by the galaxy power spectrum and bispectrum.

The paper is structured as follows: Section 2 describes the basis of perturbation theory upon which the data-vector estimators and covariance matrix terms are computed. In Section 3, we present the covariance terms. Section 4 contains the specifics of the analysis performed. In Section 5, the compression formalism is presented. In Section 6, we present the results of applying the MCMC + KL method. Section 7 describes the performance of the PCA + KL compression in obtaining constraints on the cosmological parameters. In Section 8, we comment on the added value given by jointly using the power spectrum to the bispectrum. We conclude in Section 9. All detailed derivations are deferred to Appendices A (redshift-space kernels), B (estimator definition), C (covariance matrix terms), D (compressed covariance formalism), and E (weights orthogonalization).

## 2 PERTURBATION THEORY WITH REDSHIFT-SPACE DISTORTIONS

### 2.1 Bias model

Since luminous objects like galaxies are not exact tracers of the underlying dark matter distribution, it is necessary to model their relationship. This is commonly referred to as ‘galaxy biasing’; the relationship could be linear/non-linear, deterministic/stochastic, local/non-local, and a function of scale and cosmic time. Understanding biasing is important in its own right as a probe of galaxy formation and evolution. Galaxy biasing was recognized when it was noticed that different populations of galaxies (e.g. spirals and ellipticals) have different clustering strengths.

A physical mechanism for galaxy biasing was suggested by Kaiser (1984) and developed by Bardeen et al. (1986), namely that galaxies would tend to form in peaks in the matter density distribution thus being more clustered than the underlying matter distribution. In this model, more massive (and thus rarer) tracers are naturally more highly biased. Biasing can be non-linear (McDonald & Roy 2009) and stochastic (Dekel & Lahav 1999). Galaxy biasing also evolves with redshift (Clerkin et al. 2015, and references therein), being larger at higher redshift. It is also scale dependent at small physical scales where the non-linear effects of galaxy formation are important, although there is almost no scale dependence above 20–40 Mpc  $h^{-1}$  (Manera & Gaztañaga 2011; Crocce et al. 2015). Another popular approach, known as the Halo model, is to parametrize the relationship between galaxies and the dark matter distribution by assuming that all galaxies reside in dark matter haloes or subhaloes, discussed e.g. in Tinker et al. (2010) and references therein. We note that biasing and the relationship between galaxies and haloes can also be constrained through combinations of various observables, e.g. galaxy positions and weak gravitational lensing.

Here, we adopt a biasing model which is a Taylor expansion of the galaxy density fluctuations  $\delta_g$  in terms of the matter fluctuations  $\delta_m$  (Fry & Gaztanaga 1993):

$$\delta_g = \sum_{i=0}^{\infty} \frac{b_i}{i!} \delta_{i,m}^{n.l.} \quad (1)$$

where  $\delta_{i,m}^{n.l.}$  is the non-linear matter density fluctuation.

In this work, only the first two terms of the above expansion are considered,  $b_1$  and  $b_2$ . The above bias model considers only the local relationship between  $\delta_g$  and  $\delta_m$ ; non-local bias terms may be included in a more accurate modelling (Chan, Scoccimarro & Sheth 2012; Baldauf et al. 2012; Bel, Hoffmann & Gaztañaga 2015).

### 2.2 Redshift-space formalism

The conversion of the galaxy redshifts from surveys like BOSS or DESI to proper distances is a cosmological model-dependent operation. In addition, the local gravitational field influences the peculiar velocities of galaxies producing redshift-space distortions (Kaiser 1987) which affect the observed redshift.

Heavens et al. (1998) have shown how to express the Fourier transform of the redshift-space galaxy overdensity fluctuation, measured in cosmological surveys,  $\delta_g^s(\mathbf{k})$  in relation to the real space linear matter fluctuation  $\delta_m(\mathbf{k})$

$$\begin{aligned} \delta_g^s[\mathbf{k}] = & F_s^{(1)}[\mathbf{k}] \delta_m(\mathbf{k}) \\ & + \frac{1}{(2\pi)^3} \int d^3k_1 d^3k_2 \delta_D(\mathbf{k} - \mathbf{k}_2 - \mathbf{k}_1) \\ & \times F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_m(\mathbf{k}_1) \delta_m(\mathbf{k}_2) \\ & + \frac{1}{(2\pi)^6} \int d^3k_1 d^3k_2 d^3k_3 \delta_D(\mathbf{k} - \mathbf{k}_3 - \mathbf{k}_2 - \mathbf{k}_1) \\ & \times F_s^{(3)}[\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3] \delta_m(\mathbf{k}_1) \delta_m(\mathbf{k}_2) \delta_m(\mathbf{k}_3), \end{aligned} \quad (2)$$

where the redshift-space distortion kernels  $F_s^{(1,2,3)}$  are given in Appendix A. In this work, all  $\mathbf{k}$ -vectors are described in terms of their components parallel  $k_{\parallel}$  and perpendicular  $k_{\perp}$  to the line of sight. We define  $\mu = k_{\parallel}/k$ . The galaxy redshift-space power spectrum is defined as

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle = (2\pi)^3 P_g^s(\mathbf{k}_1) \delta_D(\mathbf{k}_1 + \mathbf{k}_2). \quad (3)$$

Substituting equation (2) into equation (3) and applying Wick’s theorem assuming that the initial perturbations are Gaussian, the power spectrum is given by:

$$\begin{aligned} P_g^s(\mathbf{k}) \equiv & P_g^{s,(1)} + P_g^{s,(2)} + P_g^{s,(3)} \\ = & (b_1 + f\mu^2)^2 P^m(k) \\ & + 2 \int \frac{d^3q}{(2\pi)^3} P^m(q) P^m(|\mathbf{k} - \mathbf{q}|) (F_s^{(2)}[\mathbf{q}, \mathbf{k} - \mathbf{q}])^2 \\ & + 6 (b_1 + f\mu^2) P^m(k) \int \frac{d^3q}{(2\pi)^3} P^m(q) F_s^{(3)}[\mathbf{q}, -\mathbf{q}, \mathbf{k}], \end{aligned} \quad (4)$$

where the growth rate is defined and parametrized as  $f(z) \equiv d \ln D / d \ln a \simeq \Omega_m^{-\gamma}(z)$  where  $\Omega_m(z)$  is the mass density parameter function at a given redshift  $z$ , the growth index  $\gamma \simeq 0.55$  for a standard cosmology (Peebles 1980; Lahav et al. 1991; Linder 2005),  $D(a)$  is the growing mode of the amplitude fluctuation, and scale factor  $a = (1 + z)^{-1}$ .  $P^m(k)$  is the linear matter power spectrum defined analogously to equation (3). In this work, when considering the power spectrum data vector we stop at tree level using only  $P_g^s = P_g^{s,(1)}$ . This choice is consistent with previous analyses (Scoccimarro, Couchman & Frieman 1999; Sefusatti et al. 2006; Song, Taruya & Oka 2015; Gagrani & Samushia 2017). The error arising from excluding the one-loop terms (Jain & Bertschinger 1994; Bernardeau et al. 2002; Taruya, Koyama & Matsubara 2008; Lazanu et al. 2016) is less than 10 per cent up to  $k_{\max} = 0.2 h \text{Mpc}^{-1}$

increasing to 30 per cent for  $k_{\max} = 0.3 \text{ hMpc}^{-1}$ . An accurate modelling of the redshift-space galaxy power spectrum was introduced by Taruya, Nishimichi & Saito (2010). This translates approximately into an error on the diagonal of the power spectrum covariance matrix computed in Section 3.1 of  $\sim 1$  per cent up to  $k_{\max} = 0.2 \text{ hMpc}^{-1}$  and  $\sim 9$  per cent for  $k_{\max} = 0.3 \text{ hMpc}^{-1}$  when including cosmic variance terms. For brevity, sometimes we write  $P_g^s(k_i^\perp, k_i^\parallel) \equiv P_g^s(k_i)$ , keeping in mind that the galaxy power spectrum depends on the parallel and perpendicular components of the wave vector. Therefore,  $P_g^s$  will not depend on the second order bias parameter  $b_2$ .

Analogously, the bispectrum is defined as (Fry 1984):

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \rangle = (2\pi)^3 B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \times \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \quad (5)$$

The relation to the linear matter density power spectrum is at first order:

$$B_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2P^m(k_1)P^m(k_2)F_s^{(1)}[\mathbf{k}_1]F_s^{(1)}[\mathbf{k}_2]F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] + \text{two cyclic terms}. \quad (6)$$

For the redshift-space galaxy bispectrum, the most recent and accurate models (up to the mild non-linear regime) were introduced by Hashimoto, Rasera & Taruya (2017) and Raccanelli et al. (2013).

### 3 COVARIANCE

In this section, we summarize the covariance terms at tree level for the power spectrum and the bispectrum, including also the cross-covariance between the two. Here are written only the final analytical expressions, while we show the full derivations in Appendix C.

#### 3.1 $C^{PP}$ : power spectrum covariance matrix

The power spectrum covariance matrix is given by two terms,

$$C_P^{ss} = C_P^{PP} + C_T^{PP}, \quad (7)$$

where  $C_P^{PP}$  is proportional to the square of the power spectrum and  $C_T^{PP}$  is proportional to the trispectrum. The first term is given by:

$$\begin{aligned} C_P^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \\ &\times \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \rangle \langle \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \rangle + 1p. \\ &\approx \frac{2\pi}{V_s \bar{k}_1^\perp \Delta k^2} D_{1234}^P P_g^s(\bar{\mathbf{k}}_1)^2 = \frac{4\pi}{V_s \bar{k}_1^\perp \Delta k^2} \delta_{13}^K P_g^s(\bar{\mathbf{k}}_1)^2, \end{aligned} \quad (8)$$

where the integrals are over a cylindrical shell  $V_i^c$  centred at each  $\bar{\mathbf{k}}_i$  with integral limits on:  $\bar{k}_i^\parallel - \Delta k^\parallel/2 \leq k_i^\parallel \leq \bar{k}_i^\parallel + \Delta k^\parallel/2$ ,  $\bar{k}_i^\perp - \Delta k^\perp/2 \leq k_i^\perp \leq \bar{k}_i^\perp + \Delta k^\perp/2$  and  $0 \leq \phi \leq 2\pi$ . From the definition of the power spectrum estimator given in Appendix B, the cylindrical bins for  $\mathbf{k}_2$  and  $\mathbf{k}_4$  are centred respectively on  $\bar{\mathbf{k}}_1$  and  $\bar{\mathbf{k}}_3$ . When a Dirac delta  $\delta_D(\mathbf{k}_i + \mathbf{k}_j)$  is used to simplify one of the integrals over the cylindrical shells, it produces a Kronecker delta  $\delta_{ij}^K$  which is equal to one and not to zero only when  $k_i^\parallel = k_j^\parallel$  and  $k_i^\perp = k_j^\perp$ .

We have defined a combination of Kronecker deltas  $D_{1234}^P = \delta_{13}^K \delta_{24}^K + \delta_{14}^K \delta_{23}^K$  in order to take into account the additional permutation, which, given the initial Dirac deltas conditions, reduces to

$D_{1234}^P = 2 \times \delta_{13}^K$ .  $V_s$  is the survey volume. The second term in equation (7) is proportional to the trispectrum and is given by

$$C_T^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) \approx \frac{1}{2\pi V_s} \int_0^{2\pi} d\phi_{13} T_g^s(\bar{k}_1^\perp, \bar{k}_1^\parallel, \bar{k}_3^\perp, \bar{k}_3^\parallel, \phi_{13}). \quad (9)$$

where  $\phi_{13}$  is the difference between the azimuthal angles of  $\bar{\mathbf{k}}_1$  and  $\bar{\mathbf{k}}_3$ . The full expanded expression is written in Appendix C, while the trispectrum definition is in Appendix B.

#### 3.2 $C^{BB}$ : bispectrum covariance matrix

For the bispectrum covariance matrix, at leading order, only the diagonal term proportional to the product of three power spectra is required:

$$\begin{aligned} C_P^{BB}(\bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \bar{\mathbf{k}}_3; \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5, \bar{\mathbf{k}}_6) &= \frac{(2\pi)^5 \Delta k_3^\parallel}{V_s \bar{k}_1^\perp \bar{k}_2^\perp \bar{k}_3^\perp \Delta k^6} \Lambda_{123}^{-1} \\ &\times D_{123456} \prod_{i=1}^3 P_g^s(\bar{\mathbf{k}}_i). \end{aligned} \quad (10)$$

where  $\Lambda$  is a function defined in Appendix B which is related to the fraction of wave-vector triplets allowed by the triangle condition such that the bispectrum estimator (B5) is unbiased.  $D_{123456}$  is a shorthand notation for the sum of all the possible permutations of pairings of  $\mathbf{k}$  vectors between the first and second triplets, encoded in Kronecker deltas, e.g.  $\delta_{14}^K \delta_{25}^K \delta_{36}^K$ . This is the symmetry factor which is equal, respectively to 1, 2, and 6 in the case of scalene, isosceles, and equilateral triangles. Full computations can be found in Appendix C.

#### 3.3 $C^{BP}$ : cross-variance matrix

The cross-variance term is also given by the sum of two parts. The first part is proportional to the product between the power spectrum and bispectrum. The second part is proportional to the tetraspectrum

$$C^{BP} = C_{ml}^{BP} + C_{Te}^{BP}. \quad (11)$$

The expression of the first term as derived in Appendix C is:

$$\begin{aligned} C_{ml}^{BP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= 2 \times \frac{(2\pi)^2}{V_s V_1^c} \Lambda_{345}^{-1} (\delta_{13}^K + \delta_{14}^K + \delta_{15}^K) \\ &\times P_g^s(\bar{\mathbf{k}}_1) B_g^s(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5), \end{aligned} \quad (12)$$

Where the semicolon separates the wave vector relative to the power spectrum from the ones of the bispectrum.

The last non-vanishing term is the one proportional to the tetraspectrum  $Te_g^s$  given by the connected part of the five points correlator

$$\begin{aligned} C_{Te}^{BP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \frac{(2\pi)^2 \Delta k_5^\parallel}{V_s} \Lambda_{345}^{-1} \prod_{i=1,3,4,5} \int \frac{dV_i^c}{V_i^c} \\ &\times \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5). \end{aligned} \quad (13)$$

The definition and analytical expression for the tetraspectrum are in Appendix B. In the appendix, it is shown that at leading order the tetraspectrum is proportional to the fourth power of the linear matter power spectrum. Being this one order higher than all the terms considered in the paper,  $C_{Te}^{BP}$  has not been included in the numerical computations.



All the terms rederived above agree with the ones derived in the literature when considering redshift-space distortions (Sefusatti et al. 2006; Scoccimarro et al. 1998).

### 3.4 Shot noise contribution

All terms of the covariance matrix have been corrected in order to account for the shot noise, which is given by the average number density of galaxies. In our analysis, we consider a DESI-like ELGs sample with  $\bar{n}_g \approx 1.175 \times 10^{-3} h^3 \text{Mpc}^{-3}$  and a survey volume  $V_s \approx 12.3 h^{-3} \text{Gpc}^3$  (Levi et al. 2013; Song et al. 2015; DESI Collaboration et al. 2016). The only correction corresponds to substituting the galaxy power spectrum with:

$$P_g^s(k) \longrightarrow P_g^s(k) + \frac{1}{\bar{n}_g}. \quad (14)$$

We made the assumption that the shot noise is well approximated by a Gaussian distribution (which is reasonable if the galaxy number density is fairly high). In that case only second moments exist, and thus contributions appear only in the 2pt correlators, and those yield power spectrum contributions to the cosmic variance. If the noise were Poisson-distributed, the corrected expressions would become much more complicated (Matarrese, Verde & Heavens 1997b).

## 4 ANALYSIS SETUP

In this section, we describe the pipeline and codes for the compression and the MCMC analyses. We compressed the power spectrum and the bispectrum with respect to the set of parameters:  $\{b_1, b_2, f, \Omega_m, \Omega_b, A_s, n_s\}$  where the first two are galaxy bias parameters,  $f$  is the linear growth rate,  $\Omega_m$  and  $\Omega_b$  are the density parameters of total matter and baryonic matter,  $A_s$  is the scalar amplitude of the primordial perturbations, and  $n_s$  is the scalar spectral index. For both cases, the fiducial cosmology was fixed to  $b_1 = 1.90$ ,  $b_2 = 0.20$ ,  $f = \Omega_m(z = 0.81)^{0.55}$ ,  $\Omega_m(z = 0) = 0.307$ ,  $\Omega_b(z = 0) = 0.0482$ , and  $A_s = 2.9 \times 10^{-9}$  similar to the one used in Kitaura et al. (2016) and close to the one reported in the *Planck* 15 results (Ade et al. 2016). The redshift used is that of the effective redshift bin ( $0.6 < z < 1.0$ ) of an ELG sample of a DESI-like survey with  $z_{\text{eff}} = 0.81$ .

The  $k$ -range chosen is  $0.01 < k < 0.2 \text{Mpc}^{-1} h$ , 12 bins for both parallel (linear binning) and perpendicular (logarithmic binning) to the line-of-sight components. We have adopted a logarithmic binning for the perpendicular components of the wave vectors in order to better capture the different features at different scales. A linear binning has been used for parallel to the line of sight. For the perpendicular component,  $\Delta \log_{10} k^\perp = 0.11827$  while for the parallel one  $\Delta k^\parallel = 0.0182 \text{Mpc}^{-1} h$ . With these settings, the resulting number of configurations satisfying the triangle condition is  $N_{\text{tr}} = 1333$  for the bispectrum and  $N_{\text{pairs}} = 132$  combinations of parallel and perpendicular components for the power spectrum.

The triangle configurations for the bispectrum are generated by a 5D loop choosing first the three perpendicular components of the sides of the triangle and secondly two of the parallel ones. The third parallel component is chosen such that the final triangle satisfies the triangle condition. All sides and projections must be in the range given above.

For the MCMC sampling, we have used 64 samplers together with the same number of Xeon E5-2650 processors connected through MPI (Gabriel et al. 2004), each with 2000 burn-in steps followed by 10 000 steps for the actual posterior sampling. The PYTHON package EMCEE was used as MCMC sampler (Foreman-Mackey et al. 2013).

Uninformative flat priors have been used both in the compression code and MCMC sampler. We employed the CAMB code (Lewis, Challinor & Lasenby 2000) in order to generate the linear matter power spectrum for different cosmological parameters.

## 5 COMPRESSION FORMALISM

### 5.1 Fisher information matrix

The log-likelihood  $\mathcal{L}$  for a Gaussian probability distribution relative to an  $n$ -dimensional data vector  $\mathbf{x}$ , can be written as

$$-2\mathcal{L} = n \ln 2\pi + \ln \det \mathbf{Cov} + (\mathbf{x} - \langle \mathbf{x} \rangle)^\top \mathbf{Cov}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle), \quad (15)$$

where  $\mathbf{Cov} = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)^\top \rangle$  is the covariance matrix and  $\langle \mathbf{x} \rangle$  is the mean of the data vector.

From this quantity, the Fisher information matrix can be defined as

$$F_{ij} = - \left\langle \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right\rangle_{\theta_{\text{ML}}} \equiv - \langle \mathcal{L}_{,ij} \rangle, \quad (16)$$

which is a measure of the curvature around the maximum-likelihood point  $\theta_{\text{ML}} = (\theta_{\text{ML}}^1, \theta_{\text{ML}}^2, \dots, \theta_{\text{ML}}^m)$ , where the  $\theta$ 's are the  $m$  model parameters and the comma notation indicates the derivatives with respect to them. In the case of a Gaussian likelihood, the Fisher matrix can be expressed as

$$F_{ij} = \frac{1}{2} \text{Tr} [\mathbf{A}_i \mathbf{A}_j + \mathbf{Cov}^{-1} \mathbf{M}_{ij}], \quad (17)$$

where the matrices  $\mathbf{A}_i$  and  $\mathbf{M}_{ij}$  are defined as  $\mathbf{A}_i \equiv \mathbf{Cov}^{-1} \mathbf{Cov}_{,i}$  and  $\mathbf{M}_{ij} \equiv \langle \mathbf{x} \rangle_{,i} \langle \mathbf{x} \rangle_{,j}^\top + \langle \mathbf{x} \rangle_{,i}^\top \langle \mathbf{x} \rangle_{,j}$ .

The diagonal entries of the Fisher matrix are related to the minimum error attainable in estimating a parameter  $\theta_i$ . In particular, in the case of single parameter estimation, the minimum attainable error is  $\Delta \theta_i^{\text{min}} = 1/(F_{ii})^{1/2}$ . When more than one parameter is considered, the full Fisher matrix is needed to compute the minimum marginalized error for each parameter, which is given by  $\Delta \theta_i^{\text{min}} = (F_{ii}^{-1})^{1/2}$ . The target of the compression is to obtain a new data vector such that for each parameter  $\theta_i$ ,  $\Delta \theta_i^{\text{min}}$  is minimized. In other words, the compression algorithm will be obtained by maximizing  $F_{ii}$  for each parameter  $\theta_i$ .

### 5.2 Karhunen–Loève compression method

A general linear transformation of the data vector  $\mathbf{x}$  with a transformation matrix  $\mathbf{B}$  is given by

$$\mathbf{y} = \mathbf{B} \mathbf{x}. \quad (18)$$

The mean and the covariance matrix for  $\mathbf{y}$  become, respectively  $\langle \mathbf{y} \rangle = \mathbf{B} \langle \mathbf{x} \rangle$  and  $\mathbf{Cov}_y = \mathbf{B}^\top \mathbf{Cov}_x \mathbf{B}$ .

In the case in which only one of the linear combinations of the data is considered,  $\mathbf{B}$  has only one row,  $\mathbf{B} = \mathbf{b}^\top$ . Therefore the diagonal entries of the Fisher matrix are given by:

$$F_{ii} = \frac{1}{2} \left( \frac{\mathbf{b}^\top \mathbf{Cov}_{,i} \mathbf{b}}{\mathbf{b}^\top \mathbf{Cov} \mathbf{b}} \right)^2 + \frac{(\mathbf{b}^\top \langle \mathbf{x} \rangle_{,i})^2}{(\mathbf{b}^\top \mathbf{Cov} \mathbf{b})}, \quad (19)$$

where again the comma notation followed by the index  $i$  stands for the derivative with respect to the model parameter  $\theta_i$ . As explained by Heavens, Jimenez & Lahav (2000), who also derive weights based on minimizing the Fisher matrix diagonal elements, it is a very complex problem to find an analytical solution for  $\mathbf{b}$  from the

full equation (19). Only by ignoring then the first term of equation (19), it is possible to find an analytic solution for  $\mathbf{b}$ . This implies assuming that the derivatives of the covariance matrix with respect to model parameters are negligible when compared to the data vector ones. Recently, Heavens et al. (2017) presented a method to relax this approximation.

For both the power spectrum and bispectrum, we numerically checked that it is reasonable to assume it to be valid, by taking the ratio between the diagonal elements of  $\mathbf{Cov}_{,i}/\sqrt{\mathbf{Cov}}$  with  $\langle \mathbf{x} \rangle_{,i}$  [corresponding to the ratio of the first and second terms in equation (19), respectively]. This ratio for the bispectrum case results to be on average  $\leq 10$  per cent for the considered parameters. There are peaks for few triangle configurations ( $\leq 10$  for more than a thousand triangles) for which the ratio is around 70 per cent–80 per cent, while the minima are around 2 per cent.

By maximizing  $F_{ii}$  as described in Tegmark et al. (1997) using a Lagrange multiplier, it follows that the compressed data vector is given by a scalar

$$y_i = \langle \mathbf{x} \rangle_{,i}^T \mathbf{Cov}^{-1} \mathbf{x} \equiv \mathbf{b}^T \mathbf{x}, \quad (20)$$

where we have defined the weighting vector  $\mathbf{b} = \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}$ .

For the compression, it is acceptable to use an analytic approximation of the covariance matrix (in our case as it is described in Section 3), as any systematic error in the covariance would mostly degrade the compression; it would not bias the parameter inference. The compression enables then to use an accurate simulated covariance for the actual inference. This has the advantage of being able to compute weights for a data vector with an arbitrary large dimension (e.g. number of triangles). The Fisher matrix diagonal element can be rewritten as

$$F_{ii} = \langle \mathbf{x} \rangle_{,i}^T \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}. \quad (21)$$

In order to apply the KL compression method, one needs to choose a fiducial set of parameters at which to compute (analytically or numerically) the derivatives of the mean. In our case, the fiducial values are reported in Section 4. We compute the numerical derivatives using the five-point method for the first derivative in one dimension (Abramowitz 1974).

## 6 MCMC OF COMPRESSED DATA VECTORS

In this section, we compare the results obtained by running the MCMC algorithm for both uncompressed and compressed data vectors. The goal is to check whether it would be possible to substitute the original data vector for its compressed version. Even if this would not bring any relevant advantage in terms of speed when computing the data vector from the theoretical model, in the case of a real cosmological survey, it will be much easier and less expensive in terms of required simulations/mock catalogues to estimate the covariance matrix of the compressed data vector rather than the full data vector. Moreover, in the case of the bispectrum, a much larger number of triangles could be used for the original data vector, allowing more information to be captured, since the dimension of the compressed covariance matrix would be reduced to the number of parameters considered in the analysis. In order to use the compression, it is necessary to convert the covariance matrix for the full data vector to the one for the compressed data vector. This is shown in Appendix D.

In this paper, we consider two cases for the data vector: the galaxy bispectrum  $\mathbf{B}_g^s$  and the joint data vector  $\begin{bmatrix} \mathbf{P}_g^s, \mathbf{B}_g^s \end{bmatrix}$ . For the latter, we

include two further cases depending on whether or not we compress the power spectrum  $\mathbf{P}_g^s$ .

Fig. 2 shows the 1D and 2D marginalized posterior distributions when only the bispectrum data vector is considered. For the considered set of parameters, there is no substantial loss of information when the compressed data vector is used, even if some degeneracies are present. In this case, the compressed vector has seven elements instead of the  $\sim 1000$  triangles for the uncompressed bispectrum. The 1D and 2D posteriors have been smoothed using a Gaussian kernel density estimation procedure, for clearer visualization. Fig. 3 shows the 1D and 2D marginalized posterior distributions when we consider both the power spectrum and the bispectrum. As can be seen, there is no qualitative or quantitative difference on the posterior distributions between compressing or not compressing the power spectrum together with the bispectrum. The precise numbers can be found in the summarizing Table 1.

## 7 POSTERIOR DISTRIBUTIONS DIRECTLY FROM COMPRESSED DATA VECTORS

It is possible to compute the 1D likelihood for each linear combination  $y_{\theta_i}$  of the original data vector obtained compressing with respect to the parameter  $\theta_i$  as done in Zablocki & Dodelson (2016):

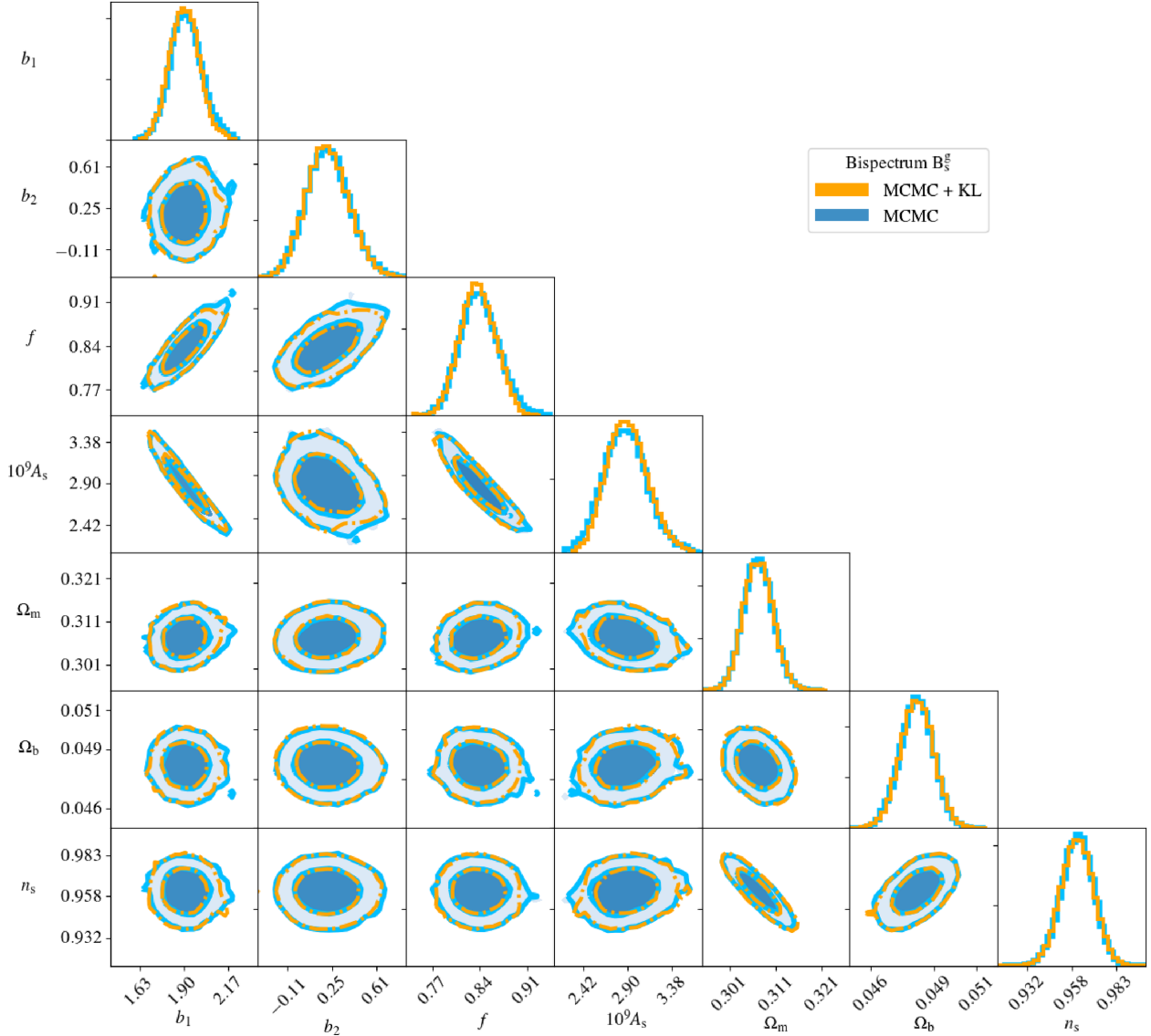
$$\ln \mathcal{L} = -\frac{(y_{\theta_i} - \bar{y}_{\theta_i})^2}{2\sigma_{(y_{\theta_i})}^2} \quad \text{with} \quad \sigma_{(y_{\theta_i})}^2 = \sum_{j=1}^{n_x} b_j^2 \mathbf{Cov}_{jj}, \quad (22)$$

where  $\bar{y}_{\theta_i} \equiv \langle \bar{y}_{\theta_i} \rangle$  is the mean of the compressed linear combination. Nevertheless, these 1D likelihoods would not be realistic since they are obtained using the compressed data-scalar  $y_{\theta_i}$  by varying only one parameter at a time. In reality  $y_{\theta_i}$  is sensitive to all the other parameters.

In order to account for this fact, we transform the set of  $y_{\theta_i}$  scalars in such a way that, at linear order, they are sensitive only to the parameter with respect to which the original data vector was compressed. One method to do so is to orthogonalize the parameter space by diagonalizing the Fisher information matrix. This has the advantage of obtaining realistic multidimensional posteriors directly from the 1D posteriors of the new set of model parameters (linear combinations of the original, physical parameters).

For comparison, we include another method presented in Zablocki & Dodelson (2016) which consists of orthogonalizing the weights in such a way that the Fisher matrix for the compressed data vector would become diagonal; we call this weights orthogonalization and it is described in Appendix E. The idea behind these two slightly different orthogonalization approaches is the same, namely to diagonalize the Fisher information matrix. The difference is that while the diagonalization in the first method is the starting point independently from compressing (or not compressing) the data vector later, in the second method the diagonalization is a consequence of the procedure used.

Both methods are approximations at linear level. Therefore as anticipated earlier they both fail (even if at different levels as it is shown later) when non-linear degeneracies are present. For example, cases where the 2D posterior distribution of a pair of parameters can no longer be approximated by an ellipse but has instead a ‘banana’-shape. This breakdown of the above procedures is in agreement with the fact that the compression method relies on the assumption that the multidimensional posteriors are Gaussian.



**Figure 2.** Bispectrum case: 1D and 2D posterior distributions for the MCMC sampling done using the full data vector (blue) and the compressed data vector obtained through the K-L compression (orange). The contours correspond to the 68 per cent and 95 per cent confidence intervals. We see that there is no substantial loss of information despite the presence of strong degeneracies between the parameters; this is quantified in Fig. 7 and Table 1. The 2D posterior distributions have been smoothed using the Gaussian kernel density estimation package provided by `scipy`.

### 7.1 Parameter space orthogonalization – PCA

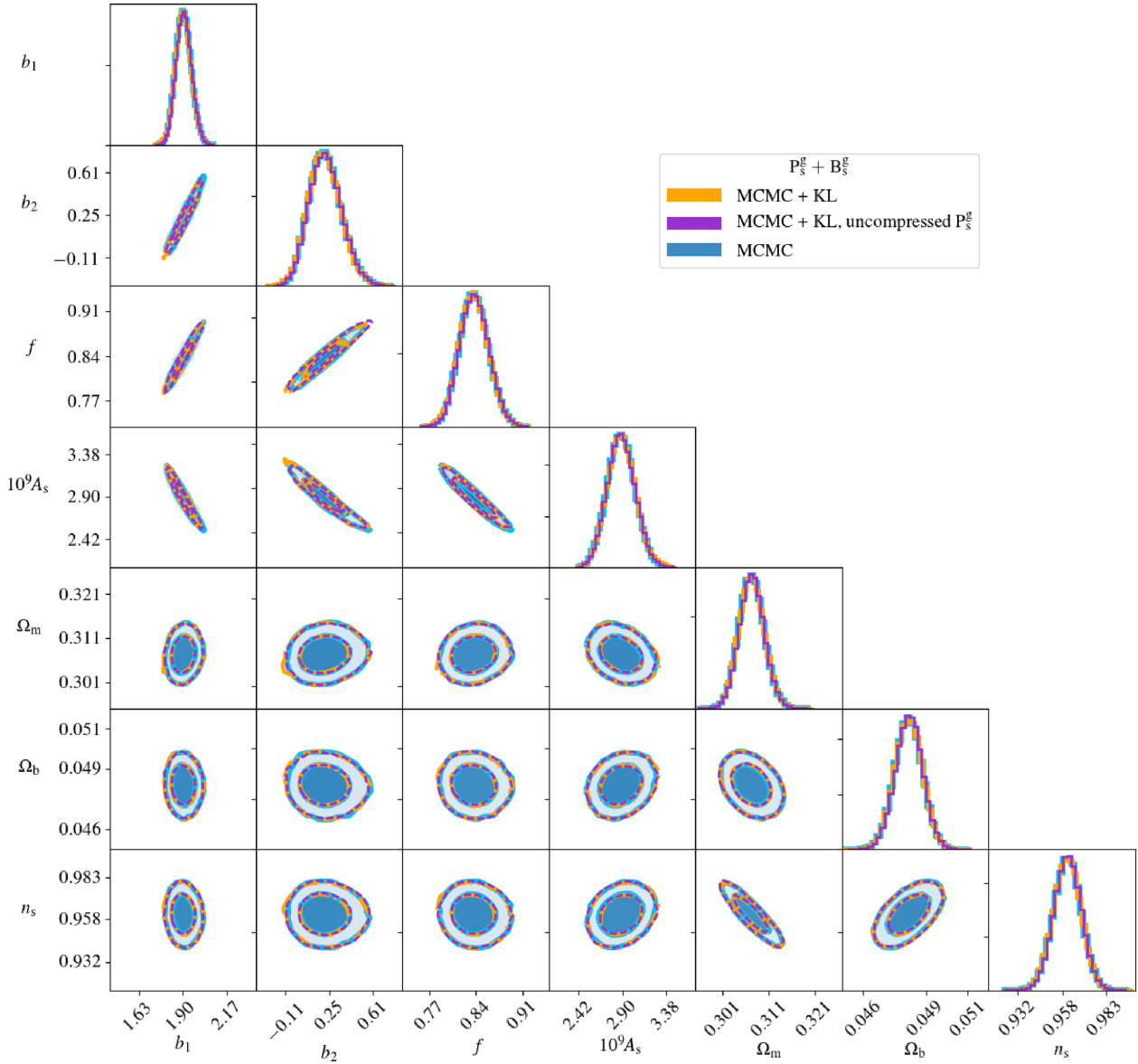
As anticipated above, the compression returns only 1D posterior distributions for each one of the parameters. Therefore, an additional step is required in order to be able to assume that these distributions correspond to 1D marginalizations from the original multidimensional distribution. This is because the KL compression with respect to one model parameter returns a linear combination of the original data vector which is still sensitive to the variation of the other model parameters. Zablocki & Dodelson (2016) obtained marginalized 1D posterior distributions by orthogonalizing the weighting vectors for all the model parameters through a Gram–Schmidt like procedure described in Appendix E. We label this method as **ORT + KL**.

In order to test whether the compression results match the ones obtained running a MCMC sampling algorithm, we also compare the 2D posterior distributions for different pairs of parameters. In the case of **ORT + KL**, the only possibility to reconstruct the multidimensional posterior is to take the outer product of the 1D marginalized posteriors. However doing so would be misleading: it would return ellipses with axes oriented at different angles to the ones given by the MCMC sampling as it can be seen in Figs 5 and 6 looking at the difference between the 2D red (**ORT + KL**) and blue (**MCMC**) contours. These axes orientation reflects the different degree of degeneracy between the parameters.

Therefore in order to avoid this difference, instead of orthogonalizing the weights we perform a PCA transformation of our parameter space before applying the KL compression. This is done by diagonalizing the Fisher information matrix using the eigenvalue decompositions

$$\mathbf{F}_{\theta_{\text{phys.}}} = \mathbf{P} \mathbf{F}_{\theta_{\text{PCA}}} \mathbf{P}^T \quad \text{where} \quad \theta_{\text{PCA}} = \mathbf{P}^T \theta_{\text{phys.}}, \quad (23)$$

where  $\mathbf{P}$  is the linear transformation matrix. After having diagonalized the Fisher matrix, we compress the data vector with respect to this new set of parameters  $\theta_{\text{PCA}}$ . The weights obtained doing so are



**Figure 3.** Joint data vector  $\left[P_g^s, B_g^s\right]$  posteriors: the colours for the 1D and 2D posterior distributions are: the full data vector (blue), the compressed data vector (orange), and the uncompressed power spectrum plus the compressed bispectrum (purple). The contours correspond to the 68 per cent and 95 per cent confidence intervals. There is no substantial loss of information due to compression, which again is quantified in Fig. 7 and Table 1.

displayed in Fig. 4 for the case of the bispectrum as data vector. The effect of a PCA decomposition is to rotate the parameter space to the axes corresponding to the degeneracies between the original set of parameters. Therefore taking the outer product of the 1D posteriors of the parameters  $\theta_{\text{PCA}}$  in order to get the multidimensional posterior distribution should return with good approximation the one sampled by the MCMC code. Once the multidimensional posterior has been reconstructed by taking the outer product of the 1D posterior distributions for the  $\theta_{\text{PCA}}$  set of parameters, this can be randomly sampled in terms of the physical parameters  $\theta_{\text{phys}}$  using the rotation matrix  $\mathbf{P}$  in order to get the 1D and 2D marginalized posterior distributions for the  $\theta_{\text{phys}}$  parameters. The results for the PCA + KL method just described can be seen looking at the green (PCA + KL) and blue (MCMC) 2D contours still in Figs 5 and 6. The PCA + KL method recovers tilted ellipses in good agreement with the MCMC ones.

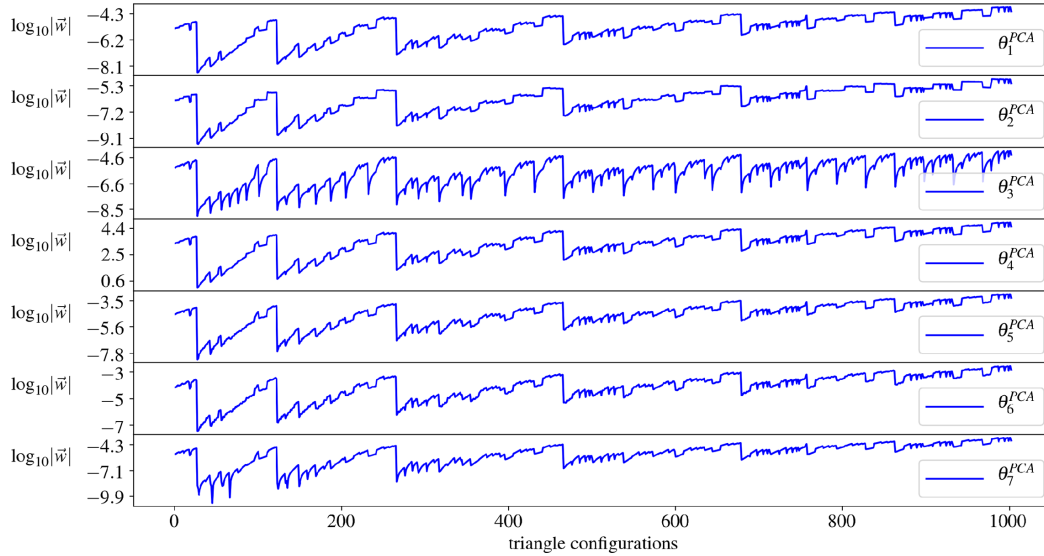
## 7.2 Comparison with MCMC sampling

Fig. 7 shows the ratio and relative difference of the 68 per cent confidence intervals of the 1D marginalized posteriors between the compression methods MCMC + KL and PCA + KL and the standard MCMC sampling. In the bispectrum case, while MCMC + KL tends to underestimate the 68 per cent confidence intervals obtained by the MCMC, PCA + KL tends to overestimate them by approximately the same amount ( $\sim 2.5$  per cent). For the joint data vector, MCMC + KL returns equivalent confidence intervals to the MCMC's ones, while PCA + KL overestimates them in average by  $\sim 2.5$  per cent (first panel) but still less than  $\sim 1$  per cent in terms of relative difference (second panel). We consider the averages of these ratios since at the same time the compression methods overestimate the 68 per cent confidence intervals for some parameters while underestimating them for others. In terms of the individual parameters



**Table 1.** The 68 per cent confidence intervals of the 1D posteriors for the bispectrum and the joint data vector  $[P_g^s, B_g^s]$  obtained from the full data vectors and the two possible compression applications, KL + MCMC sampling or PCA + KL compression. Since the marginalized 1D distributions are not perfectly Gaussian (symmetric with respect to the maximum-likelihood point) in the table is reported for each parameter, the halved difference between the maximum and minimum values included in the 68 per cent confidence interval. The four rows in the third panel show the improvement in percentage on the constraints gained by using the joint data vector instead of the bispectrum alone for all the compression methods considered. The percentage shows indicates how much smaller are the constraints obtained using  $[P_g^s, B_g^s]$  as data vector than the ones computed using just  $B_g^s$ . In the last panel, the percentages are relative to the improvement obtained by adding the bispectrum to the power spectrum. The added value of the bispectrum with respect to power spectrum alone consists in 68 per cent confidence intervals  $\sim 4$  times smaller for the amplitude-like parameters  $b_1, f$ , and  $A_s$ .

		$\Delta\theta = \frac{\theta_{\max}^{68\text{ per cent}} - \theta_{\min}^{68\text{ per cent}}}{2}$						
		$\Delta b_1$	$\Delta b_2$	$\Delta f$	$\Delta 10^9 A_s$	$\Delta \Omega_m$	$\Delta \Omega_b$	$\Delta n_s$
$P_g^s$	MCMC	0.251	—	0.111	0.856	0.0041	0.0009	0.0116
	MCMC + KL	0.095	0.174	0.028	0.232	0.0031	0.0008	0.0091
$B_g^s$	MCMC + KL	0.090	0.172	0.027	0.220	0.0032	0.0008	0.0093
	PCA + KL	0.101	0.173	0.029	0.249	0.0032	0.0008	0.0092
$[P_g^s, B_g^s]$	MCMC	0.047	0.132	0.022	0.151	0.0027	0.0007	0.0079
	MCMC + KL	0.047	0.134	0.022	0.153	0.0027	0.0007	0.0079
	' ' ( $P_g^s$ uncomp.)	0.048	0.134	0.022	0.152	0.0027	0.0006	0.0078
	PCA + KL	0.049	0.136	0.022	0.157	0.0027	0.0007	0.0079
per cent $(\Delta\theta^B - \Delta\theta^{P+B}) / \Delta\theta^B$	MCMC	51%	24%	23%	35%	13%	15%	13%
	MCMC + KL	47%	22%	17%	31%	13%	14%	15%
	' ' ( $P_g^s$ uncomp.)	47%	22%	17%	31%	14%	17%	16%
	PCA + KL	52%	22%	22%	37%	14%	15%	14%
per cent $(\Delta\theta^P - \Delta\theta^{P+B}) / \Delta\theta^P$	MCMC	81%	—%	80%	82%	34%	27%	32%
	MCMC + KL	81%	—%	80%	82%	33%	27%	31%
	' ' ( $P_g^s$ uncomp.)	81%	—%	80%	82%	34%	29%	33%
	PCA + KL	81%	—%	80%	82%	33%	27%	32%



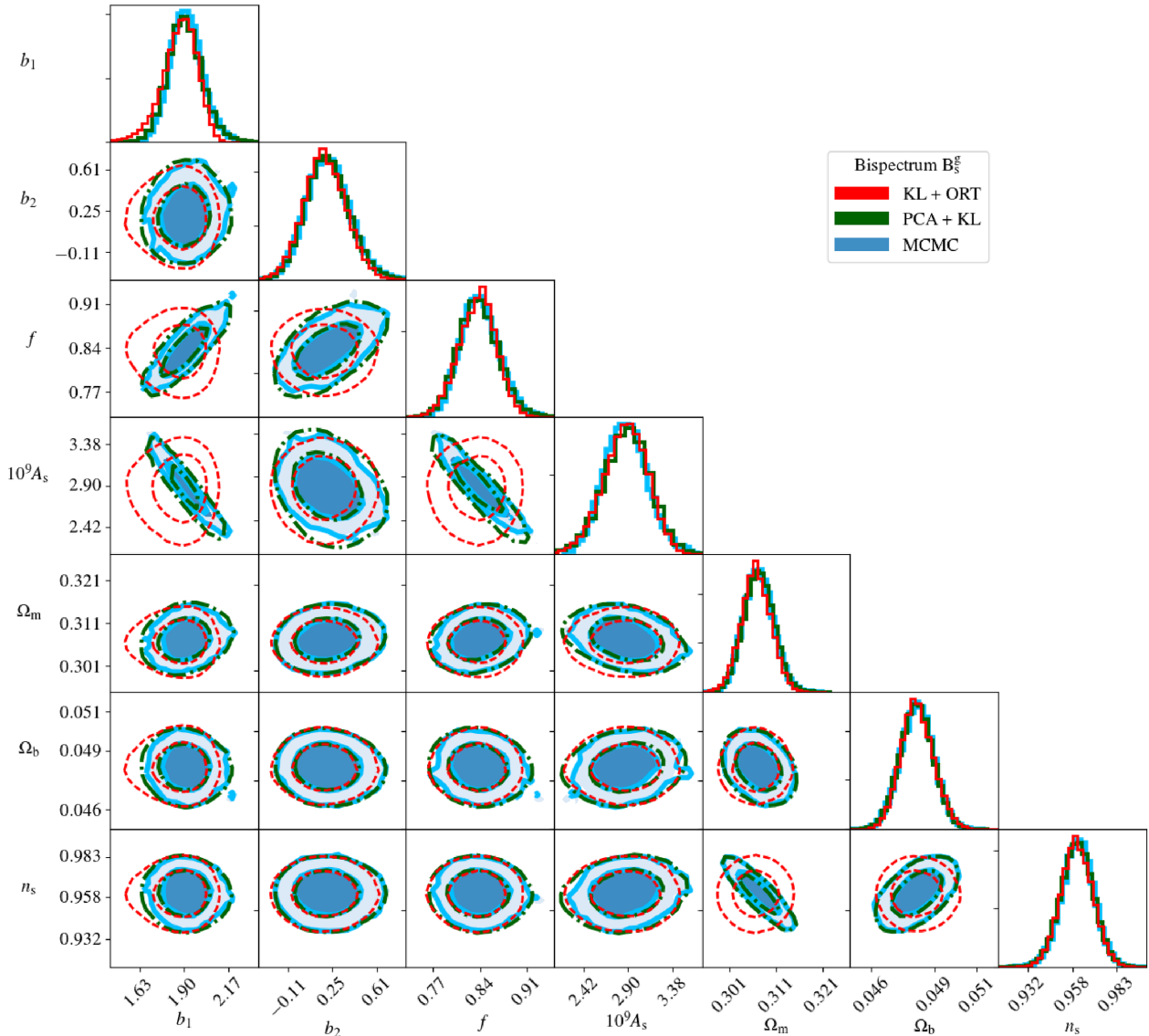
**Figure 4.** Logarithm of the absolute value of the weights for all the triangle configurations used in the bispectrum data vector. Each row corresponds to the weights for the bispectrum  $B_g^s$  with respect to a specific linear combination  $\theta_{PCA}$  of the original cosmological parameters obtained by diagonalizing the Fisher information matrix. The discontinuities observed reflect the five loops used to produce the set of triangles. From left to right, the average size of the sides of the triangles increases, from the smallest triangle up to the largest. The amplitude of the weights slightly depends on the size of the triangle (increasing from left to right). The fluctuations are far more influenced by the shape and orientation of the triangles (which can be seen within each loop).

68 per cent confidence intervals, PCA + KL diverges at most from the MCMC's ones respectively by  $\sim 7$  per cent and  $\sim 6$  per cent in the case of the bispectrum and the joint data vector. MCMC + KL diverges at most, respectively, by  $\sim 6$  per cent and  $\sim 4$  per cent in the case of the bispectrum and the joint data vector.

Both in the case of the bispectrum and joint data vector  $[P_g^s, B_g^s]$ , the compression for both MCMC + KL (Figs 2 and 3) and PCA +

KL (Figs 5 and 6) methods well matches the 1-2D contours derived from the MCMC sampling. This shows that in the bispectrum case there is no relevant difference between MCMC + KL and PCA + KL methods while there is a very small one in the case of the joint data vector.

In the bispectrum case (Fig. 5) when the 2D posterior distributions are considered, only the contours derived by orthogonalizing the parameter space before the compression (PCA + KL) have elliptical



**Figure 5.** Comparison between the MCMC-derived posteriors and the ones obtained using only the compression. The 1D and 2D posterior distributions are relative to the data vector  $B_g^s$ , for the three different cases: MCMC (blue), KL + ORT compression (red), and PCA + KL compression (green). The contours correspond to the 68 per cent and 95 per cent confidence intervals (Fig. 7 and Table 1 for numerical values).

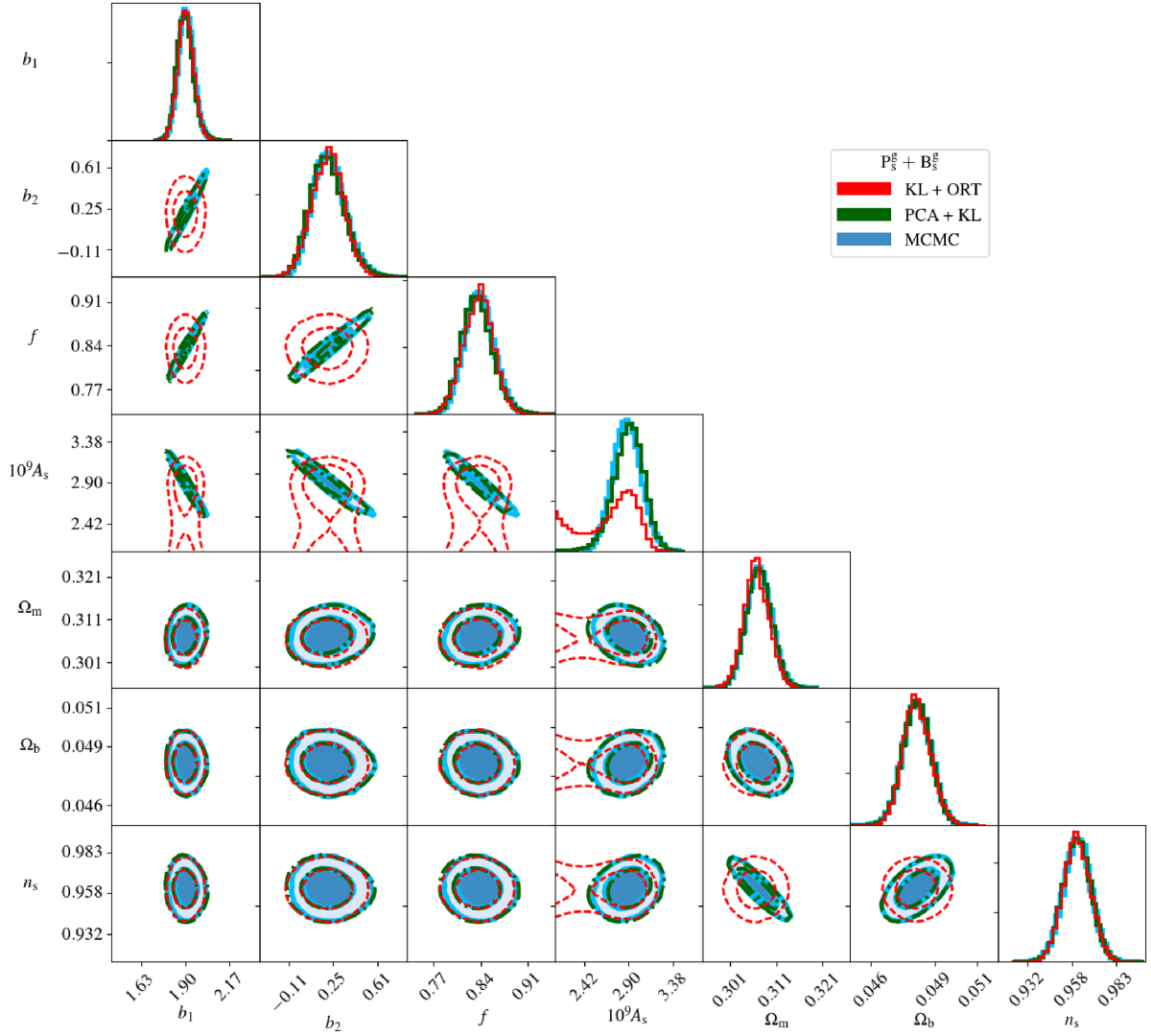
shapes with the right inclination. In the case of the joint data vector instead (Fig. 6), the KL method alone fails to recover even the 1D posterior for the parameter  $A_s$ , producing an unphysical secondary peak, while using PCA + KL returns 1D and 2D contours equivalent to the MCMC's ones. This shows that the orthogonalization of the parameter space (PCA step) performs better than the weights orthogonalization for combinations of parameters presenting strong non-linear degeneracies. As described in the following subsection, the compression methods PCA + KL and KL + ORT breakdown when the  $k_{\max}$  used drops below a certain threshold value. For the KL + ORT method, this is  $\sim 0.2 \text{ Mpc}^{-1} h$  as shown in Fig. 6, while for PCA + KL, the threshold value is lower. Applying the compression up to those mildly non-linear scales on real data like the BOSS DR12 requires to modify the model in order to take into account non-local bias terms. However, this would not require the introduction of additional bias parameters. This is because assuming that even if the galaxy bias is non-local in Eulerian space, it is in Lagrangian space, and hence the additional non-local bias terms

can be related at first order to the linear bias  $b_1$  as done in Gil-Marín et al. (2015).

The parameter set used in our analysis has been chosen to have strong degeneracies in order to test the applicability of the MCMC + KL and PCA + KL methods. When working with real data, the parameter set is usually designed to be less degenerate, for example considering only  $\sigma_8$  instead of  $\Omega_m$  and  $A_s$  or also combining  $\sigma_8$  with  $b_1$ ,  $b_2$ , and  $f$ .

In absence of non-linear (banana shape) degeneracies, the compression PCA + KL can be a valid and much faster substitute to the standard MCMC sampling since it returns realistic multidimensional posterior distributions. In particular, compression could be used to accurately forecast the constraints for different sets of parameters. It would also be possible to qualitatively study the degeneracies present in each set.

Running both compression algorithms (KL or PCA) takes approximately 20 min for seven parameters, with the time depending on the number of intervals for each parameter range and



**Figure 6.** 1D and 2D posterior distributions using as data vector  $\begin{bmatrix} P_g^s & B_g^s \end{bmatrix}$ , for the three different cases: MCMC (blue), KL + ORT compression (red), and PCA + KL compression (green). The contours correspond to the 68 per cent and 95 per cent confidence intervals (Fig. 7 and Table 1 for numerical values). The KL method fails to reproduce the correct contours, both 1D and 2D, since for the used value of  $k_{\max}$  it fails to work, producing an unphysical secondary peak for the parameter  $A_s$ . On the contrary, the PCA + KL method recovers with very good agreement the MCMC contours.

therefore how many times the linear matter power spectrum is computed. As a rough estimate, if we consider 100 intervals for each of the seven parameter ranges,  $P^m$  will be recomputed 300 times (no need to recompute the matter power spectrum when varying only one of the parameters  $b_1$ ,  $b_2$ ,  $f$ , or  $A_s$ ). For a standard MCMC with 64 samplers and  $10^4$  steps instead of 300 times  $P^m$  needs to be recomputed  $64 \times 10^4$  times. Therefore, the compression is more than  $\sim 2100$  times faster than the MCMC.

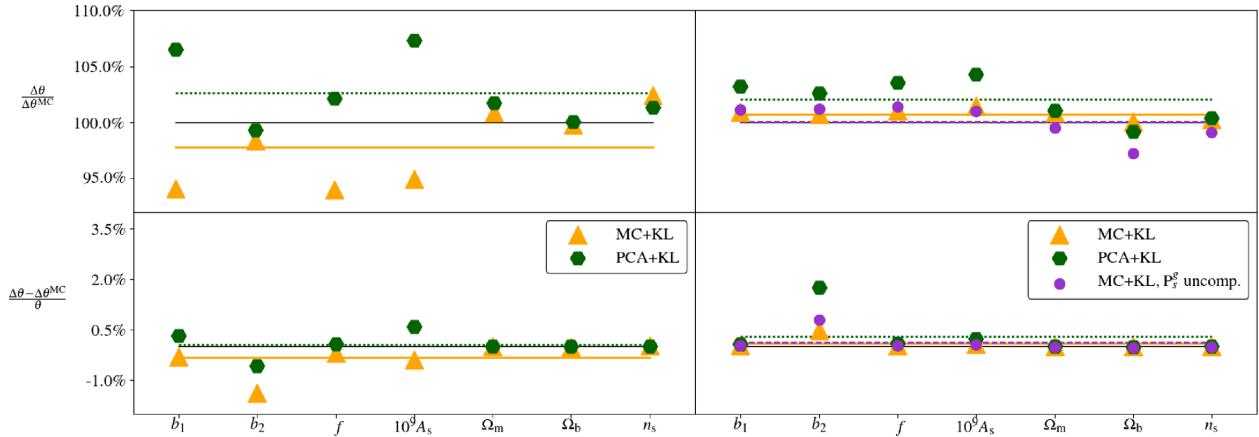
The compression PCA + KL is also much less demanding in terms of computing resources. It can be run in few minutes on a laptop with a single Intel i7 processor while for the MCMC we have used 64 threads working in parallel on 64 Xeon E5-2650 processors connected using MPI for  $\sim 72$  h.

### 7.3 Limitations of the compression

Unfortunately, the orthogonalization prescriptions do not work when ‘strong’ non-linear degeneracies are present in the param-

eter space and the Fisher matrix can no longer be diagonalized in practice. This happens because the multidimensional posterior distribution can no longer be approximated by a multivariate Gaussian distribution. In our work, this happens when  $k_{\max}$  is lowered, reducing the information accessible through the power spectrum and the bispectrum.

This failure of the method manifests itself with the appearance of unphysical secondary peaks in the posterior distributions. For example, when the MCMC (or MCMC + KL) returns 2D posterior contours with degeneracies that are banana-shaped, like in the case of the joint data vector  $\begin{bmatrix} P_g^s & B_g^s \end{bmatrix}$  for the parameters  $(b_1, b_2, f, \sigma_8)$ , the compression alone (KL + ORT or PCA + KL) fails to recover the 1D and 2D posteriors. In these cases, both the KL + ORT and PCA + KL produce unphysical secondary peaks in the posterior distribution. However, as shown in Fig. 6, PCA + KL still works considering only larger scales ( $k_{\max}$  lower), where standard perturbation theory gives more accurate predictions, than when using only the KL + ORT method.



**Figure 7.** Ratio (first panel) and relative difference (second panel) between the 68 per cent confidence intervals of the 1D marginalized posteriors for all the cosmological parameters obtained using either the compression plus the MCMC sampling (MCMC + KL, orange triangles) or just the compression (PCA + KL, green hexagons) with the ones obtained running the MCMC on the full data vector. The black lines in all plots represent the reference values given by running a MCMC sampling using the full data vector. The ratios  $\Delta\theta/\Delta\theta^{\text{MC}}$  give an idea of how much the 68 per cent confidence intervals obtained through compression differ from the one given when using the full data vector. The relative difference  $(\Delta\theta - \Delta\theta^{\text{MC}})/\theta$  is helpful, since it scales the difference between the compressed and full data-vector results with respect to the value of the chosen parameters. The first column shows the results when just the bispectrum  $B_g^s$  is considered, where the largest discrepancy happens for the parameters most degenerate between each other's ( $b_1, f, A_s$ ). Using the MCMC + KL method on the bispectrum proves to be negligible with loss of information. In the second column, the ratios and relative differences in the case of the joint data vector  $[P_g^s + B_g^s]$  are shown. In all the subplots, the horizontal lines show the averages of the ratios and relative differences of the same colour. For  $[P_g^s + B_g^s]$ , the compression is optimal, both using MCMC + KL and PCA + KL. Compressing or not the power spectrum together with the bispectrum seems to produce no relevant statistical difference. The bottom right subplot  $b_2$  appears to be the parameters whose divergence from the MCMC result is greater with respect to its fiducial value.

In the case of the bispectrum, the breakdown of the PCA + KL method happens for  $k_{\text{max}} < 0.18 \text{ Mpc}^{-1}h$  while for the joint data vector  $[P_g^s, B_g^s]$  for  $k_{\text{max}} < 0.17 \text{ Mpc}^{-1}h$ . A potential solution to this limitation could be the application of the Gaussianization method proposed by Schuhmann, Joachimi & Peiris (2016) which we plan to include in a future work. This failure also happens when  $V_s$  or  $n_g$  are one order of magnitude smaller. However, these scenarios are below the specifications of current and future cosmological surveys. Another solution is to consider less parameters by either fixing to a fiducial value one or more of the most degenerate ones or by rewriting them in terms of combinations which absorb the degeneracies (e.g.  $f\sigma_8$ ). It is important to note that in this paper we have on purpose considered an extreme case of degenerate parameter space in order to test the method applicability. In a realistic case, a much less degenerate parameter space is usually considered when analysing data. In the considered parameter space, the main cause of parameters degeneracy is due to having both  $A_s$ ,  $\Omega_m$ , and  $f$  instead of using only  $\sigma_8$  and  $f$ . It is then reasonable to assume that the method would still work when a more complex biasing model is considered by adding a further bias parameter  $b_3$  or the tidal bias  $b_{s2}$ , provided that  $A_s$  and  $\Omega_m$  are substituted with  $\sigma_8$ . A more accurate biasing model would be needed in an application to real data in order to describe the bias non-local nature observed in simulations (Sheth, Chan & Scoccimarro 2013; Modi, Castorina & Seljak 2017).

## 8 JOINT DATA-VECTOR ADDED VALUE

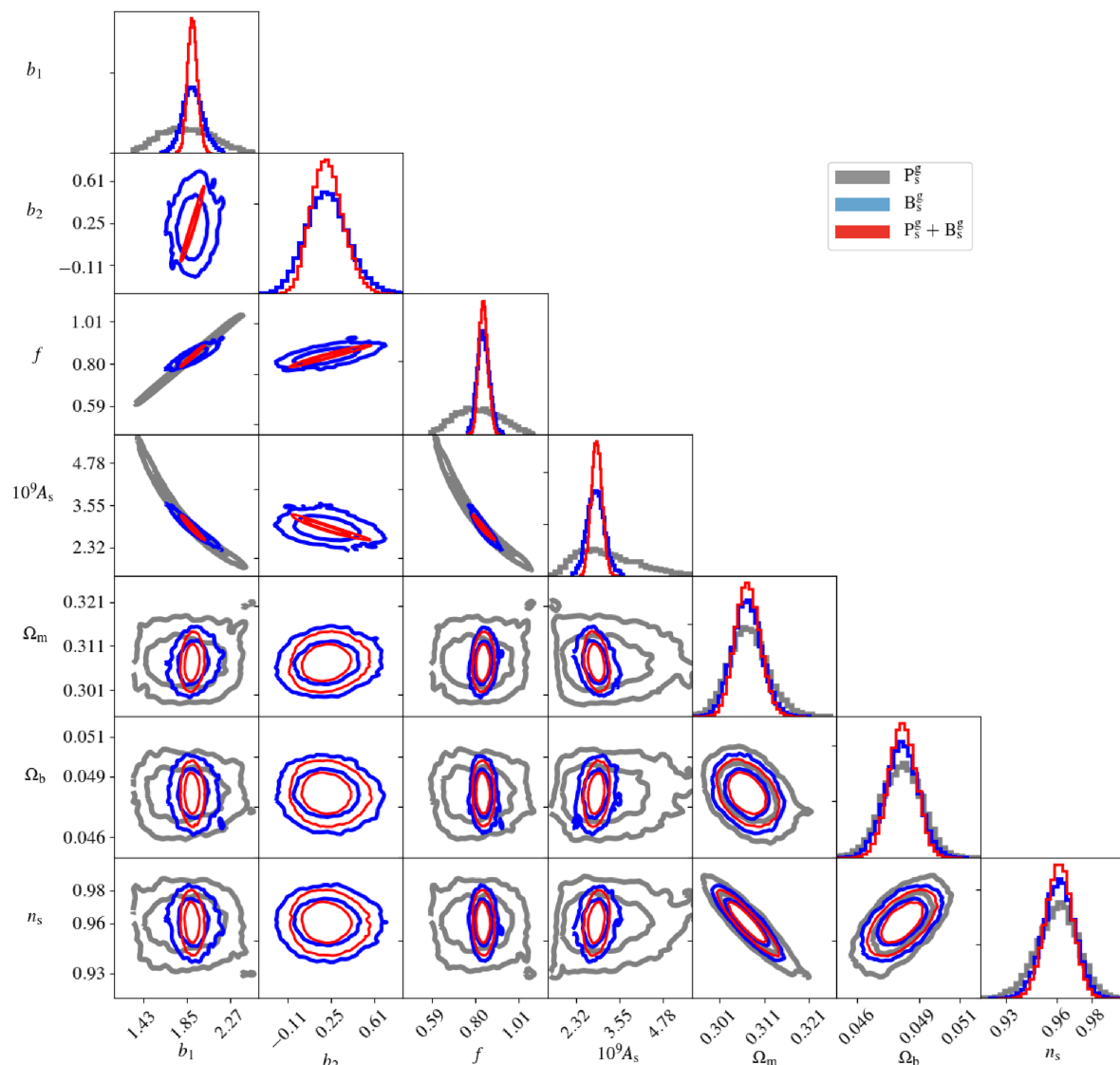
In both cases presented in this paper, either running an MCMC sampling on the compressed data vector or doing the parameter estimation directly from the compression (MCMC + KL or PCA + KL), the added value of combining the power spectrum with the bispectrum is qualitatively (Fig. 8) and quantitatively (Fig. 7 and Table 1) evident. This improvement obtained using the joint data

vector is particularly relevant for all the degenerate amplitude-like parameters:  $b_1$ ,  $b_2$ ,  $f$ , and  $A_s$ . For the redshift bin of a DESI-like survey considered in our analysis, the improvement for the above-mentioned parameters obtained using the joint data vector with respect using only the bispectrum is respectively of 52 per cent, 22 per cent, 22 per cent, and 37 per cent (using PCA + KL). The improvement with respect to the power spectrum alone is even greater, in particular: 81 per cent, 80 per cent, and 82 per cent for  $b_1$ ,  $f$ , and  $A_s$ .

Comparing our results to Sefusatti et al. (2006) and in particular their table 8 for  $k_{\text{max}} = 0.2 \text{ Mpc}^{-1}h$ , it is possible to see a similar effect due to including the bispectrum in the analysis together with the power spectrum. The similarity consists in the fact that the parameters that more benefit from this addition are the bias ones ( $b_1$ ,  $b_2$ ) and the amplitude ones ( $A_s$ ,  $\sigma_8$ ). The greater improvement that we find for certain parameters by adding the bispectrum to the power spectrum with respect to Sefusatti et al. can be explained by the fact that we consider redshift-space distortions, which are encoded in the growth rate parameter  $f$ . This increases the degeneracies already present between  $b_1$ ,  $b_2$ ,  $\sigma_8$ ,  $A_s$ , and  $\Omega_m$ . Therefore, the added value of the bispectrum in lifting the degeneracies increases. Together with this, in Sefusatti et al.'s paper the covariance matrix of the bispectrum includes also off-diagonal terms, which decrease the constraining power of the bispectrum since these terms describe the correlation between different triangle configurations.

For what concerns, the most recent bispectrum measurements from BOSS DR12 data by Gil-Marín et al. (2016), it is important to point out that the data vector considered is given by the power spectrum monopole and quadrupoles together with the bispectrum monopole. Using the bispectrum monopole implies washing out some of its information when integrating over the azimuthal and polar angles. This could explain the lower impact of the bispectrum monopole with respect to the power spectrum monopole and quadrupole in constraining the parameters.





**Figure 8.** Forecasted posteriors for the redshift bin of  $q$  DESI-like survey for what concerns the combined two- and three points statistics. 1D and 2D contours are derived from MCMC sampling using for the three data vectors:  $P_g^s$  (grey),  $B_g^s$  (blue), and  $[P_g^s + B_g^s]$  (red). Combining two and three points statistics enables to drastically reduce and break the degeneracies in the parameter space, in particular the ones between  $b_1$ ,  $A_s$ , and  $b_2$ . This improvement obtained using the joint data vector is therefore particularly evident for all these amplitude-like parameters (Table 1). In particular, it is qualitatively evident from this plot the benefit of adding the bispectrum to the power spectrum. It is important to notice that for the adopted model the galaxy power spectrum does not depend on the parameter  $b_2$ . Even with one parameter less to constrain than the bispectrum, as expected, the MCMC for the galaxy power spectrum alone does not well converge for most of the degenerate parameters.

Moreover in this case, the covariance matrix used to do parameter inference is estimated from galaxy mocks, therefore it includes not only off-diagonal terms describing the correlation between different triangles, but also the noise due to the fact that it is an estimated quantity.

Finally, in the recent paper by Byun et al. (2017) similar improvement to ours have been obtained in parameter constraints (table 3), in particular for what concerns  $b_1$  and  $\sigma_8$  (proxy for  $A_s$ ).

The efficiency of the compression PCA + KL also drastically improves if one considers the joint data vector  $[P_g^s, B_g^s]$  as can be seen in Figs 3 and 6. Indeed, the combination of power spectrum and bispectrum reduces the degeneracies between the considered parameters and the results obtained by running approximately 20 min long compression pipeline almost perfectly match the ones given

by a three days of MCMC sampling run in parallel on 64 processors (right-hand panel, Fig. 7).

As is well known in the literature, the degeneracy between the bias parameters (in particular at the linear order  $b_1$ ) with the amplitude of the dark matter perturbations  $\sigma_8$  or the primordial perturbations scalar amplitude  $A_s$  cannot be broken using only the power spectrum. These degeneracies are even larger when redshift-space distortions are considered. On the other hand, the bispectrum alone can (in theory) lift these degeneracies, even if it requires including the quadratic bias parameter  $b_2$ . In any case being a 3pt statistic, it is more difficult to measure and analyse from real surveys than 2pt statistics. Therefore, combining power spectrum and bispectrum is of fundamental importance in order to obtain the best possible constraints, especially in light of the large data sets that

are going to become available in the near future (DESI, *Euclid*, PFS, etc.).

## 9 CONCLUSIONS

In this paper, we have shown that it is possible to compress the information about cosmological parameters contained in the galaxy power spectrum and bispectrum and to extract this information in an efficient way. In particular, we have shown:

(i) Compressing the data vector using the K-L algorithm before running an MCMC sampling gives negligible loss of information in terms of parameters constraints (Figs 2, 3, and 7). In particular, running an MCMC sampling on the compressed data vector (MCMC + KL) returns 68 per cent confidence intervals less than 1 per cent larger in terms of relative difference than the MCMC's ones. This happens in both the cases of the bispectrum and power spectrum plus bispectrum ( $B_g^s$  and  $[P_g^s, B_g^s]$ ). For real surveys, this would allow us to drastically reduce the number of simulations needed to numerically estimate the covariance matrix. This is because the dimension of the compressed covariance matrix corresponds to the number of model parameters, not the number of the original data-vector elements.

(ii) Orthogonalizing the parameter space through the diagonalization of the Fisher matrix before applying the compression (PCA + KL), proves to be competitive with the MCMC sampling with minimal loss of constraining power (Figs 5 and 6). PCA + KL returns for both  $B_g^s$  and  $[P_g^s, B_g^s]$  68 per cent confidence intervals less than 1 per cent different in terms of relative difference from the ones obtained running the MCMC sampling for the full data vectors. This method cannot be applied when the parameter space presents non-linear degeneracies, since a multidimensional Gaussian posterior distribution is no longer a valid approximation. Using the joint data vector  $[P_g^s, B_g^s]$  lowers the minimum  $k_{\max}$  necessary for the only-compression method (PCA + KL) to work to  $k_{\max} \simeq 0.17 \text{ Mpc}^{-1}h$  (while in case of just the bispectrum, this threshold is  $k_{\max} \simeq 0.18 \text{ Mpc}^{-1}h$ ). These values have been obtained considering the redshift bin of a DESI-like survey. Moreover these values can be further lowered if the parameter space we considered is reduced to a less degenerate one, for example  $(b_1, b_2, f, \sigma_8)$ .

(iii) Byun et al. (2017) reduced the covariance matrix dimension without significant loss of information by using a proxy that aggregates the matter bispectrum over a subset of Fourier configurations. In particular, they obtained their best results in terms of constraints on  $\Lambda$  cold dark matter parameters from combining the power spectrum with the modal decomposition of the bispectrum (Fergusson, Regan & Shellard 2012; Regan et al. 2012). Their results demonstrate that the modal bispectrum performs as well as the Fourier bispectrum, even with considerably fewer modes (10) than Fourier configurations (95).

The main difference with the approach presented in this work is that we use the original full galaxy bispectrum data vector in order to compress it. This does not have any limitations in terms of original size of the bispectrum data vector. Therefore, it allows us to access the full information content achievable through the bispectrum. Moreover, the number of elements of the compressed data vector in our case is independent of the number of elements of the original data vector. In our case, we considered  $\sim 1000$  triangle configurations and *seven* parameters which give a compression of at least two orders of magnitude. In Byun et al. (2017), it is left

for future work to check whether the achieved compression of the bispectrum data vector by an order of magnitude would improve if more triangle configurations are taken into consideration.

(iv) As already quantified using real data sets like BOSS (Gil-Marín et al. 2016), the use of the joint data vector  $[P_g^s, B_g^s]$  significantly increases the constraining power on the cosmological parameters compared to using only the power spectrum or bispectrum individually. In the case of an ELGs sample for a DESI-like survey, the improvement obtained by combining power spectrum and bispectrum is quantified and can be visualized in the second panel of Fig. 7 and the second and third rows of Table 1. The constraints obtained considering power spectrum and bispectrum together are up to 52 per cent smaller than the constraints obtained using only the bispectrum. The difference is even greater with respect to the power spectrum alone. Together with lifting the degeneracies between amplitude-like parameters as it can be seen in Fig. 8, the 68 per cent confidence intervals of the marginalized 1D posteriors for the joint data vector are up to  $\sim 5$  times smaller than the power spectrum ones.

(v) Using the compression PCA + KL as analysis method is much faster than MCMC and less computationally demanding (few minutes on a single processor compared to days using several processors working in parallel). Since it is relatively easy to implement, it can be used to ‘sample’ different sets of parameters and obtain reliable constraints for a given model without having to wait days for each one of them as in the case of the MCMC.

(vi) The pre-compression PCA transformation allows us to better capture the nature of the degeneracies between the chosen parameters, returning realistic multidimensional posterior distributions that follow closely the MCMC ones. Also in the case of ‘strong’ degeneracies, orthogonalizing the parameter space before compressing returns contours qualitatively more realistic and closer to the MCMC than orthogonalizing the weights after compressing.

Future work will include the study of how the compression method applied here is affected by the choice of the set of considered parameters. Finally, we would like to test the methods presented here using BOSS data or simulations for the upcoming DESI, *Euclid*, and PFS surveys. Using MCMC + KL or PCA + KL compression methods has the potential of becoming a standard fast and reliable approach to adopt when dealing with large data vector as in the case of higher order statistics.

## ACKNOWLEDGEMENTS

The authors thank the anonymous referee for the useful comments. DG is grateful to William Jennings and Keir Rogers for carefully reading the manuscript. DG is supported by the Perren and the IMPACT funds. MM acknowledges funding from STFC Consolidated Grants RG84196 and RG70655 LEAG/506 and has received funding from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 6655919. BJ acknowledges support by an STFC Ernest Rutherford Fellowship, grant reference ST/J004421/1. OL acknowledges support from a European Research Council Advanced Grant FP7/291329. c (Kernighan 1988) and PYTHON 2.7 (Rossum 1995) have been used together with many packages like IPYTHON (Perez & Granger 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), SCIPY (Jones et al. 2001), and MATPLOTLIB (Hunter 2007).

## REFERENCES

- Abramowitz M., 1974, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover Publications, Incorporated, New York
- Ade P. A. R. et al., 2014, *A&A*, 571, A16
- Ade P. A. R. et al., 2016, *A&A*, 594, A13
- Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, 86, 083540
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
- Bel J., Hoffmann K., Gaztañaga E., 2015, *MNRAS*, 453, 259
- Bernardeau F., 1994, *ApJ*, 433, 1
- Bernardeau F., Brax P., 2011, *J. Cosmol. Astropart. Phys.*, 1106, 019
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
- Borisov A., Jain B., 2009, *Phys. Rev. D*, 79, 103506
- Byun J., Eggemeier A., Regan D., Seery D., Smith R. E., 2017, *MNRAS*, 471, 1581
- Chan K. C., Scoccimarro R., Sheth R. K., 2012, *Phys. Rev. D*, 85, 083509
- Clerkin L., Kirk D., Lahav O., Abdalla F. B., Gaztañaga E., 2015, *MNRAS*, 448, 1389
- Crocce M., Castander F. J., Gaztañaga E., Fosalba P., Carretero J., 2015, *MNRAS*, 453, 1513
- Dekel A., Lahav O., 1999, *ApJ*, 520, 24
- DESI Collaboration et al., 2016, preprint ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
- Eisenstein D. J. et al., 2011, *AJ*, 142, 72
- Fergusson J. R., Regan D. M., Shellard E. P. S., 2012, *Phys. Rev. D*, 86, 063511
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Fry J. N., 1984, *ApJ*, 279, 499
- Fry J. N., Gaztanaga E., 1993, *ApJ*, 413, 447
- Fry J. N., Scherrer R. J., 1994, *ApJ*, 429, 36
- Fry J. N., Seldner M., 1982, *ApJ*, 259, 474
- Gabriel E. et al., 2004, Open MPI: goals, concept, and design of a next generation MPI implementation, in *Proceedings of the 11th European PVM/MPI Users' Group Meeting*. Budapest, Hungary, p.97
- Gagran P., Samushia L., 2017, *MNRAS*, 467, 928
- Gangui A., Lucchin F., Matarrese S., Mollerach S., 1994, *ApJ*, 430, 447
- Gil-Marín H., Noreña J., Verde L., Percival W. J., Wagner C., Manera M., Schneider D. P., 2015, *MNRAS*, 451, 539
- Gil-Marín H., Percival W. J., Verde L., Brownstein J. R., Chuang C.-H., Kitaura F.-S., Rodríguez-Torres S. A., Olmstead M. D., 2016, *MNRAS*, 465, 1757
- Groth E. J., Peebles P. J. E., 1977, *ApJ*, 217, 385
- Hartlap J., Simon P., Schneider P., 2007, *A&A*, 464, 399
- Hashimoto I., Rasera Y., Taruya A., 2017, *Phys. Rev. D*, 96, 043526
- Heavens A. F., Matarrese S., Verde L., 1998, *MNRAS*, 301, 797
- Heavens A. F., Jimenez R., Lahav O., 2000, *MNRAS*, 317, 965
- Heavens A. F., Sellentin E., de Mijolla D., Vianello A., 2017, *MNRAS*, 472, 4244
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jain B., Bertschinger E., 1994, *ApJ*, 431, 495
- Joachimi B., Shi X., Schneider P., 2009, *A&A*, 508, 1193
- Jones E. et al., 2001, SciPy: Open source scientific tools for Python. Available at: <http://www.scipy.org/>
- Juszkiewicz R., Hellwing W. A., van de Weygaert R., 2013, *MNRAS*, 429, 1206
- Kaiser N., 1984, *ApJ*, 284, L9
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kernighan B. W., 1988, *The C Programming Language*, 2nd edn. Prentice Hall, Upper Saddle River, NJ
- Kitaura F.-S. et al., 2016, *MNRAS*, 456, 4156
- Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, *MNRAS*, 251, 128
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Lazanu A., Giannantonio T., Schmittfull M., Shellard E. P. S., 2016, *Phys. Rev. D*, 93, 083517
- Levi M. et al., 2013, preprint ([arXiv:1308.0847](https://arxiv.org/abs/1308.0847))
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Liguori M., Sefusatti E., Fergusson J. R., Shellard E. P. S., 2010, *Adv. Astron.*, 2010, 980523
- Linder E. V., 2005, *Phys. Rev. D*, 72, 043529
- Manera M., Gaztañaga E., 2011, *MNRAS*, 415, 383
- Matarrese S., Verde L., Heavens A. F., 1997a, *MNRAS*, 290, 651
- Matarrese S., Verde L., Heavens A. F., 1997b, *MNRAS*, 290, 651
- McDonald P., Roy A., 2009, *J. Cosmol. Astropart. Phys.*, 8, 020
- Modi C., Castorina E., Seljak U., 2017, *MNRAS*, 472, 3959
- Mohammed I., Seljak U., Vlah Z., 2017, *MNRAS*, 466, 780
- Peebles P., 1980, *The Large-scale Structure of the Universe*. Princeton Series in Physics, Princeton Univ. Press, Princeton, NJ
- Peebles P. J. E., Groth E. J., 1975, *ApJ*, 196, 1
- Perez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Pielorz J., Rödiger J., Tereno I., Schneider P., 2010, *A&A*, 514, A79
- Raccanelli A. et al., 2013, *MNRAS*, 436, 89
- Regan D. M., Schmittfull M. M., Shellard E. P. S., Fergusson J. R., 2012, *Phys. Rev. D*, 86, 123524
- Rossum G., 1995, *Technical Report, Python Reference Manual*. Amsterdam, The Netherlands
- Schmittfull M. M., Regan D. M., Shellard E. P. S., 2013, *Phys. Rev. D*, 88, 063512
- Schmittfull M., Baldauf T., Seljak U., 2015, *Phys. Rev. D*, 91, 043530
- Schuhmann R. L., Joachimi B., Peiris H. V., 2016, *MNRAS*, 459, 1916
- Scoccimarro R., 2000, *ApJ*, 544, 597
- Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, *ApJ*, 496, 586
- Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, *ApJ*, 517, 531
- Sefusatti E., Crocce M., Pueblas S., Scoccimarro R., 2006, *Phys. Rev. D*, 74, 023522
- Sheth R. K., Chan K. C., Scoccimarro R., 2013, *Phys. Rev. D*, 87, 083002
- Slepian Z. et al., 2015, preprint ([arXiv:1512.02231](https://arxiv.org/abs/1512.02231))
- Slepian Z. et al., 2017, *MNRAS*, 468, 1070
- Song Y.-S., Taruya A., Oka A., 2015, *J. Cosmol. Astropart. Phys.*, 1508, 007
- Takada M. et al., 2014, *PASJ*, 66, R1
- Taruya A., Koyama K., Matsubara T., 2008, *Phys. Rev. D*, 78, 123534
- Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
- Taylor A., Joachimi B., 2014, *MNRAS*, 442, 2728
- Tegmark M., Taylor A., Heavens A., 1997, *ApJ*, 480, 22
- Tellarini M., Ross A. J., Tasinato G., Wands D., 2016, *J. Cosmol. Astropart. Phys.*, 6, 014
- The Dark Energy Survey Collaboration 2005, preprint ([arXiv:astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- van der Walt S., Colbert S. C., Varoquaux G., 2011, preprint ([arXiv:1102.1523](https://arxiv.org/abs/1102.1523))
- Verde L., Heavens A. F., Matarrese S., Moscardini L., 1998, *MNRAS*, 300, 747
- Verde L., Wang L., Heavens A. F., Kamionkowski M., 2000, *MNRAS*, 313, 141
- Zablocki A., Dodelson S., 2016, *Phys. Rev. D*, 93, 083525

**APPENDIX A: REDSHIFT-SPACE KERNELS**

The redshift-space kernels derived from standard perturbation theory as reported in Matarrese et al. (1997a) and used in the expression for the power spectrum and bispectrum in this work are given by:

$$\begin{aligned}
 F_s^{(1)}[k] &= b_1 + f\mu^2; \\
 F_s^{(2)}[k_1, k_2] &= b_1 J_s^{(2)}[k_1, k_2] + f\mu^2 K_s^{(2)}[k_1, k_2] + \frac{1}{2}b_2 + \frac{b_1 f}{2} \left[ \mu_1^2 + \mu_2^2 + \mu_1 \mu_2 \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \right] \\
 &\quad + f^2 \left[ \mu_1^2 \mu_2^2 + \frac{\mu_1 \mu_2}{2} \left( \mu_1^2 \frac{k_1}{k_2} + \mu_2^2 \frac{k_2}{k_1} \right) \right]; \\
 F_s^{(3)}[k_1, k_2, k_3] &= b_1 J_s^{(3)}[k_1, k_2, k_3] + f\mu^2 K_s^{(3)}[k_1, k_2, k_3] + \frac{b_2}{2} f\mu_3^2 + \frac{b_3}{6} + \frac{b_2}{2} f\mu_1 \mu_2 \frac{k_2}{k_1} + \frac{b_2}{2} f\mu_1 \mu_3 \frac{k_3}{k_1} \\
 &\quad + b_1 f^2 \mu_2^2 \mu_3^2 + 2b_1 f^2 \mu_1 \mu_2 \mu_3^2 \frac{k_1}{k_2} + b_1 f^2 \mu_2 \mu_3^3 \frac{k_3}{k_2} + \frac{b_1}{2} \mu_1^2 \mu_2 \mu_3 \frac{k_1^2}{k_2 k_3} + f^3 \mu_1^2 \mu_2^2 \mu_3^2 + 3f^3 \mu_1 \mu_2^2 \mu_3^3 \frac{k_3}{k_1} + \frac{1}{2} f^3 \mu_1 \mu_2 \mu_3^4 \frac{k_3^2}{k_1 k_2} \\
 &\quad + J_s^{(2)}[k_2, k_3] \left( b_2 + b_1 f\mu_1^2 + b_1 f\mu_1 \mu_{2+3} \frac{k_{2+3}}{k_1} \right) \\
 &\quad + K_s^{(2)}[k_2, k_3] \left( b_1 f\mu_{2+3}^2 + b_1 f\mu_1 \mu_{2+3} \frac{k_1}{k_{2+3}} + 2f^2 \mu_1^2 \mu_{2+3}^2 + f^2 \mu_1 \mu_{2+3}^3 \frac{k_{2+3}}{k_1} + f^2 \mu_1^3 \mu_{2+3} \frac{k_1}{k_{2+3}} \right), \tag{A1}
 \end{aligned}$$

where the  $K$ 's and  $J$ 's are given by:

$$\begin{aligned}
 J_s^{(2)}[k_1, k_2] &= \frac{5}{7} + \frac{k_1 \cdot k_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \left( \frac{k_1 \cdot k_2}{k_1 k_2} \right)^2; \\
 K_s^{(2)}[k_1, k_2] &= \frac{3}{7} + \frac{k_1 \cdot k_2}{2k_1 k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{4}{7} \left( \frac{k_1 \cdot k_2}{k_1 k_2} \right)^2; \\
 J_s^{(3)}[k_1, k_2, k_3] &= J_s^{(2)}[k_2, k_3] \left[ \frac{1}{3} + \frac{1}{3} \frac{k_1 \cdot (k_2 + k_3)}{(k_2 + k_3)^2} + \frac{4}{9} \frac{k \cdot k_1}{k_1^2} \frac{k \cdot (k_2 + k_3)}{(k_2 + k_3)^2} \right] \\
 &\quad - \frac{2}{9} \frac{k \cdot k_1}{k_1^2} \frac{k \cdot (k_2 + k_3)}{(k_2 + k_3)^2} \frac{k_3 \cdot (k_2 + k_3)}{k_3^2} + \frac{1}{9} \frac{k \cdot k_2}{k_2^2} \frac{k \cdot k_3}{k_3^2}; \tag{A2}
 \end{aligned}$$

$$K_s^{(3)}[k_1, k_2, k_3] = 3J_s^{(3)}[k_1, k_2, k_3] - \frac{k \cdot k_1}{k_1^2} J_s^{(2)}[k_2, k_3] - \frac{k \cdot (k_1 + k_2)}{(k_1 + k_2)^2} K_s^{(2)}[k_1, k_2]. \tag{A3}$$

In the above expressions, we defined  $\mu = k_1^\parallel / k_1$ ,  $k = k_1 + k_2 + k_3$  and  $\mu_{2+3} \equiv (k_2^\parallel + k_3^\parallel) / |k_2 + k_3|$ .

**APPENDIX B: ESTIMATORS DEFINITION AND UNBIASEDNESS CHECK**

In this appendix, we present the definition of the power spectrum and the bispectrum estimators, as well as the definition of the trispectrum and the tetraspectrum, which are necessary for the computation of the full covariance matrix for the joint data vector  $[P_g^s, B_g^s]$ .

Including the effect of redshift-space distortions in our analysis requires enlarging the number of parameters needed to describe a particular configuration of the bispectrum. In this case, it is possible to see that, considering the distant observer approximation, which consists in assuming that all the line-of-sight vectors paired to each wave vector are parallel between each others, the natural symmetry to exploit is the cylindrical one. In other words, what characterizes the redshift effect on the bispectrum are the parallel to the line-of-sight components of the wave vectors, from now on labelled as  $k^\parallel$ . Therefore, the natural set of coordinates to describe each wave vector is the cylindrical one:  $(k^\parallel, k^\perp, \phi)$  where  $k^\perp$  and  $\phi$  describe the component of the wave vector laying on the perpendicular plane to the line of sight. Hence, for what concerns the bispectrum in redshift space, from the original 9 degrees of freedom, three are cancelled by translational invariance given by the closed triangle condition  $\delta_D(k_1 + k_2 + k_3)$ .

Moreover, from the chosen coordinates above it is possible to see that there is a further symmetry which is the rotation along the line of sight of a particular triangle of wave vectors. Different configurations given simply by rotating the same triangle around the line of sight give the same value for the bispectrum. The coordinate  $\phi$  describes these rotations and expresses this symmetry. As a consequence of the symmetries mentioned, we are left with five remaining degrees of freedom, describing all possible configurations, which are given by:  $\{k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel\}$ , from which we can derive all the other quantities (e.g.  $k_3^\parallel = -k_1^\parallel - k_2^\parallel$ ). It is necessary then to define a new type of bin for our estimator: a cylindrical annuli defined by  $\phi \in [0, 2\pi]$ ,  $\bar{k}^\perp \in [\bar{k}^\perp - \Delta k^\perp/2, \bar{k}^\perp + \Delta k^\perp/2]$  and  $\bar{k}^\parallel \in [\bar{k}^\parallel - \Delta k^\parallel/2, \bar{k}^\parallel + \Delta k^\parallel/2]$ . For thin annuli, the surface area is given by  $A_c = 2\pi \Delta k \bar{k}^\perp$  and hence the volume is given by  $V^c = 2\pi \Delta k^2 \bar{k}^\perp$ . The 3D Dirac's delta can be decomposed as

$$\delta_D(k_1 + k_2 + k_3) = \delta_D(k_1^\parallel + k_2^\parallel + k_3^\parallel) \delta_D(k_1^\perp + k_2^\perp + k_3^\perp). \tag{B1}$$



## B1 Power spectrum

A standard way to define an estimator for the power spectrum is the following (Peebles 1980):

$$\hat{P}_g^s(\bar{\mathbf{k}}_1) = \frac{1}{V_s} \int_{V_1^c} \frac{dV_1^c}{V_1^c} \int_{V_2^c} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2), \quad (B2)$$

where  $V_s$  is the survey volume. Notice that both cylindrical bins are centred on  $\bar{\mathbf{k}}_1$  and therefore by definition  $\bar{\mathbf{k}}_1 = \bar{\mathbf{k}}_2$ . We specify that the power spectrum for the redshift galaxy field depends on the wave vector, precisely on its perpendicular and parallel components to the line of sight, and not on its module. Following the definition given in equation (B2), it is straightforward to check whether the estimator is unbiased; one only needs to take the average

$$\begin{aligned} \langle \hat{P}_g^s(\bar{\mathbf{k}}_1) \rangle &= \frac{1}{V_s} \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle \\ &= \frac{1}{V_s} \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 (2\pi)^3 P_g^s(\mathbf{k}_1) \\ &= \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} \int_{V^c(\mathbf{k}_1)} dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P_g^s(\mathbf{k}_1) = \int_{V^c(\mathbf{k}_1)} \frac{dV_1^c}{V^c(\mathbf{k}_1)} P_g^s(\mathbf{k}_1) \\ &= \int_{\bar{k}_1^\perp - \frac{\Delta k}{2}}^{\bar{k}_1^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_1^\parallel - \frac{\Delta k}{2}}^{\bar{k}_1^\parallel + \frac{\Delta k}{2}} \frac{dk_1^\perp dk_1^\parallel k_1^\perp}{\bar{k}_1^\perp \Delta k^2} P_g^s(\mathbf{k}_1) \approx P_g^s(\bar{\mathbf{k}}_1). \end{aligned} \quad (B3)$$

where the standard expression for the galaxy power spectrum at leading order is given by

$$P_g^s(\mathbf{k}_1) = P^m(k_1) F_s^{(1)}(\mathbf{k}_1)^2. \quad (B4)$$

## B2 Bispectrum estimator

Proceeding now as shown in Joachimi, Shi & Schneider (2009), we define the following estimator for the bispectrum

$$\hat{B}_g^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) = \frac{(2\pi)^2 \Delta k_3^\parallel}{V} \Lambda^{-1}(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \prod_{i=1}^3 \int_{V_i^c} \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3), \quad (B5)$$

where  $\Lambda$  is a function related to the fraction of wave-vectors triplets allowed by the triangle condition and defined such that the bispectrum estimator (B5) is unbiased, which can be checked by taking the average of the estimator, which also is reported in the appendix. The expression for  $\Lambda$  is

$$\Lambda(k_1^\perp, k_2^\perp, k_3^\perp) = 2\pi \int_0^\infty dr^\perp r^\perp \prod_{i=1}^3 J_0(k_i^\perp r^\perp) = \frac{4}{\sqrt{2k_1^{\perp 2} k_2^{\perp 2} + 2k_1^{\perp 2} k_3^{\perp 2} + 2k_2^{\perp 2} k_3^{\perp 2} - k_1^{\perp 4} - k_2^{\perp 4} - k_3^{\perp 4}}}, \quad (B6)$$

if  $|k_1^\perp - k_2^\perp| < k_3^\perp < k_1^\perp + k_2^\perp$  or 0 otherwise.  $J_0$  is the zeroth-order spherical Bessel function. It is also possible to check that for the bispectrum, the estimator is unbiased

$$\begin{aligned} \langle \hat{B}_g^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \rangle &= \frac{(2\pi)^2 \Delta k_3^\parallel}{V_s} \Lambda^{-1}(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel) \int_{V_1^c} \int_{V_2^c} \int_{V_3^c} (2\pi)^3 (\delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3))^2 B_g^s(k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel) \\ &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \int_{V_1^c} \int_{V_2^c} \int_{V_3^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{g,123}^s, \end{aligned} \quad (B7)$$

where a shorthand notation for the bispectrum has been introduced and where we have used the approximation  $\delta_D^2 \simeq V_s / (2\pi)^3 \delta_D$  from Joachimi et al. (2009). Since the bispectrum is invariant under rotation around the line of sight, we integrate now the angular part, namely over  $\phi$

$$\begin{aligned} \int_0^{2\pi} d\phi_1 \int_0^{2\pi} d\phi_2 \int_0^{2\pi} d\phi_3 \delta_D(\mathbf{k}_1^\perp + \mathbf{k}_2^\perp + \mathbf{k}_3^\perp) &= \int d\phi_1 d\phi_2 d\phi_3 \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} e^{i(k_1^\perp + k_2^\perp + k_3^\perp) r^\perp} \\ &= \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} \int_0^{2\pi} d\phi_1 e^{ik_1^\perp r^\perp \cos(\phi_1 - \phi_r)} \times \int_0^{2\pi} d\phi_2 e^{ik_2^\perp r^\perp \cos(\phi_2 - \phi_r)} \int_0^{2\pi} d\phi_3 e^{ik_3^\perp r^\perp \cos(\phi_3 - \phi_r)} \\ &= \int \frac{d^2 \mathbf{r}^\perp}{(2\pi)^2} (2\pi)^3 J_0(k_1^\perp r^\perp) J_0(k_2^\perp r^\perp) J_0(k_3^\perp r^\perp) = (2\pi)^2 \int_0^\infty dr^\perp r^\perp \prod_{i=1}^3 J_0(k_i^\perp r^\perp) \\ &= 2\pi \Lambda(k_1^\perp, k_2^\perp, k_3^\perp), \end{aligned} \quad (B8)$$

from which it is possible to see that in this case  $\Lambda$  depends only on the perpendicular components of the wave vectors. This agrees with the fact that on the orthogonal plane, the wave-vectors components must form a closed triangle. In one of the steps above we used the argument

that integrating a trigonometric function over an interval of  $2\pi$  gives the same result no matter what are the extrema of integration as long as the interval is of  $2\pi$ . For completeness, the Bessel functions of order 0 are defined as

$$J_0(x) = \int_0^{2\pi} \frac{d\phi}{2\pi} e^{ix \cos \phi}. \quad (\text{B9})$$

Then, inserting the expression for  $\Lambda$  (B8) into equation (B7), it is possible to verify that our estimator is unbiased

$$\begin{aligned} \langle \hat{\mathbf{B}}_{\mathbf{g},123}^s \rangle &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \prod_{i=1}^3 \int_{V_{\text{ic}}} \frac{d^2 \mathbf{k}_i^\perp d k_i^\parallel}{2\pi \bar{k}_i^\perp \Delta k^2} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \mathbf{B}_{\mathbf{g}}^s(k_1^\perp, k_2^\perp, k_3^\perp, k_1^\parallel, k_2^\parallel) \\ &= (2\pi)^2 \Delta k_3^\parallel \Lambda_{123}^{-1} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_i^\parallel - \frac{\Delta k}{2}}^{\bar{k}_i^\parallel + \frac{\Delta k}{2}} \frac{d k_i^\perp d k_i^\parallel}{2\pi \bar{k}_i^\perp \Delta k^2} k_i^\perp, 2\pi \Lambda_{123} \delta_D(k_1^\parallel + k_2^\parallel + k_3^\parallel) \mathbf{B}_{\mathbf{g},123}^s \\ &\approx \delta_{123}^K \mathbf{B}_{\mathbf{g}}^s(\bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel), \end{aligned} \quad (\text{B10})$$

where  $\delta^K$  is a Kronecker delta. In the last step, the thin-shell approximation has been used in order to bring out the bispectrum from the integrals. The standard expression for the galaxy bispectrum is given by

$$\mathbf{B}_{\mathbf{g}}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2 \mathbf{P}^m(k_1) \mathbf{P}^m(k_2) \mathbf{F}_s^{(1)}(\mathbf{k}_1) \mathbf{F}_s^{(1)}(\mathbf{k}_2) \mathbf{F}_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] + 2 \text{p.} \quad (\text{B11})$$

### B3 Trispectrum definition

The trispectrum is defined as

$$\langle \delta_{\mathbf{g}}^s(\mathbf{k}_1) \delta_{\mathbf{g}}^s(\mathbf{k}_2) \delta_{\mathbf{g}}^s(\mathbf{k}_3) \delta_{\mathbf{g}}^s(\mathbf{k}_4) \rangle_c = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathbf{T}_{\mathbf{g}}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (\text{B12})$$

where the subscript ‘c’ indicates that trispectrum is the connected part of the four points correlation function. Therefore, the leading order terms are of two types:  $\mathbf{T}_{\mathbf{g}}^{s(2)}$  and  $\mathbf{T}_{\mathbf{g}}^{s(3)}$  respectively characterized by the RSD perturbation kernels  $\mathbf{F}_s^{(2)}$  and  $\mathbf{F}_s^{(3)}$ . Starting from the first one, we have that:

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathbf{T}_{\mathbf{g}}^{s(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \langle \mathbf{F}_s^{(1)}[\mathbf{k}_1] \mathbf{F}_s^{(1)}[\mathbf{k}_2] \delta_{k_1} \delta_{k_2} \frac{1}{(2\pi)^3} \int d^3 k_a d^3 k_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \delta_{k_a} \delta_{k_b} \\ &\times \int d^3 k_c d^3 k_d \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \mathbf{F}_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \delta_{k_c} \delta_{k_d} \rangle + 5 \text{p.} \\ &= \frac{1}{(2\pi)^6} \mathbf{F}_s^{(1)}[\mathbf{k}_1] \mathbf{F}_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \int d^3 k_c d^3 k_d \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \mathbf{F}_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \langle \delta_{k_1} \delta_{k_2} \delta_{k_a} \delta_{k_b} \delta_{k_c} \delta_{k_d} \rangle + 5 \text{p.} \end{aligned} \quad (\text{B13})$$

From the last line of the previous expression it follows that, when Wick’s theorem is applied, there are three different ways to pair the wave vectors.

$$\mathbf{T}_{\mathbf{g}}^{s(2)} \begin{cases} \mathbf{T}_{\mathbf{g}}^{s(2a)} & \Leftrightarrow \langle \delta_{k_1} \delta_{k_2} \rangle \langle \delta_{k_a} \delta_{k_c} \rangle \langle \delta_{k_b} \delta_{k_d} \rangle \times 2 \quad (\text{sym.}) \\ \mathbf{T}_{\mathbf{g}}^{s(2b)} & \Leftrightarrow \langle \delta_{k_1} \delta_{k_a} \rangle \langle \delta_{k_2} \delta_{k_c} \rangle \langle \delta_{k_b} \delta_{k_d} \rangle \times 4 \quad (\text{sym.}) \\ \mathbf{T}_{\mathbf{g}}^{s(2c)} & \Leftrightarrow \langle \delta_{k_1} \delta_{k_c} \rangle \langle \delta_{k_2} \delta_{k_a} \rangle \langle \delta_{k_b} \delta_{k_d} \rangle \times 4 \quad (\text{sym.}) \end{cases}$$

The first term represents a non-connected one-loop correction to the power spectrum covariance matrix (Fry 1984; Mohammed, Seljak & Vlah 2017). For completeness, we just show below that the simplified expression makes explicit the fact that it is an unconnected term of the full four points correlator

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \mathbf{T}_{\mathbf{g}}^{s(2a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) &= \frac{2}{(2\pi)^6} \mathbf{F}_s^{(1)}[\mathbf{k}_1] \mathbf{F}_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c d^3 k_d \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \mathbf{F}_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \\ &\times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_a + \mathbf{k}_c) \delta_D(\mathbf{k}_b + \mathbf{k}_d) \mathbf{P}_{k_1}^m \mathbf{P}_{k_a}^m \mathbf{P}_{k_b}^m + 5 \text{p.} \\ &= 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \mathbf{P}_{\mathbf{g}}^s(\mathbf{k}_1) \int d^3 k_a d^3 k_b \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \mathbf{F}_s^{(2)}[-\mathbf{k}_a, -\mathbf{k}_b] \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \mathbf{P}_{k_a}^m \mathbf{P}_{k_b}^m + 5 \text{p.} \\ &= 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \mathbf{P}_{\mathbf{g}}^s(\mathbf{k}_1) \int d^3 k_a \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_3 - \mathbf{k}_a]^2 \mathbf{P}_{k_a}^m \mathbf{P}_{|\mathbf{k}_3 - \mathbf{k}_a|}^m + 5 \text{p.} \\ &= 2(2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \mathbf{P}_{\mathbf{g}}^s(\mathbf{k}_1) \int d^3 k_a \mathbf{F}_s^{(2)}[\mathbf{k}_a, \mathbf{k}_3 - \mathbf{k}_a]^2 \mathbf{P}_{k_a}^m \mathbf{P}_{|\mathbf{k}_3 - \mathbf{k}_a|}^m + 5 \text{p.} \end{aligned} \quad (\text{B14})$$

We then look at the first connected tree level term  $\mathbf{T}_{\mathbf{g}}^{s(2b)}$

$$\begin{aligned}
& (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
&= \frac{4}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d^3 k_a d^3 k_b d^3 k_c d^3 k_d \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \\
&\quad \times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_c) \delta_D(\mathbf{k}_b + \mathbf{k}_d) P_{k_1}^m P_{k_a}^m P_{k_b}^m + 5 p. \\
&= 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] P_{k_1}^m P_{k_2}^m \int d^3 k_b d^3 k_d \delta_D(\mathbf{k}_b - \mathbf{k}_1 - \mathbf{k}_3) \delta_D(\mathbf{k}_d - \mathbf{k}_2 - \mathbf{k}_4) \\
&\quad \times F_s^{(2)}[-\mathbf{k}_1, \mathbf{k}_1 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_4] P_{k_b}^m \delta_D(\mathbf{k}_b + \mathbf{k}_d) + 5 p. \\
&= 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[-\mathbf{k}_1, \mathbf{k}_1 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_2, -\mathbf{k}_1 - \mathbf{k}_3] \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) P_{k_1}^m P_{k_2}^m P_{|\mathbf{k}_1 + \mathbf{k}_3|}^m + 5 p.. \tag{B15}
\end{aligned}$$

from which it is possible to directly write the third term which result to be exactly the same

$$\begin{aligned}
& (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(2c)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
&= 4 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_1, -\mathbf{k}_2 - \mathbf{k}_3] \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) P_{k_1}^m P_{k_2}^m P_{|\mathbf{k}_2 + \mathbf{k}_3|}^m + 5 p.. \tag{B16}
\end{aligned}$$

Finally, it is possible to consider the last term  $T_g^{s(3)}$

$$\begin{aligned}
& (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
&= \langle F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \delta_{k_1} \delta_{k_2} \delta_{k_3} \frac{1}{(2\pi)^6} \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \delta_{k_a} \delta_{k_b} \delta_{k_c} \rangle + 3 p. \\
&= \frac{1}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \langle \delta_{k_1} \delta_{k_2} \delta_{k_3} \delta_{k_a} \delta_{k_b} \delta_{k_c} \rangle + 3 p.. \tag{B17}
\end{aligned}$$

Like we did before, we apply now the Wick's theorem. In this case, there are two different ways to pair the wave vectors.

$$T_g^{s(3)} \begin{cases} T_g^{s(3a)} & \Leftrightarrow (\langle \delta_{k_1} \delta_{k_2} \rangle \langle \delta_{k_3} \delta_{k_a} \rangle \langle \delta_{k_b} \delta_{k_c} \rangle + 2 p.) \times 3 (\text{sym.}) \\ T_g^{s(3b)} & \Leftrightarrow \langle \delta_{k_1} \delta_{k_a} \rangle \langle \delta_{k_2} \delta_{k_b} \rangle \langle \delta_{k_3} \delta_{k_c} \rangle \times 6 (\text{sym.}) \end{cases}$$

The first option represents again a non-connected one-loop correction to the power spectrum covariance matrix.

$$\begin{aligned}
& (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
&= \frac{3}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
&\quad \times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_1}^m P_{k_3}^m P_{k_b}^m + 2 p. \\
&= 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
&\quad \times \delta_D(\mathbf{k}_3 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_b}^m + 2 p. \\
&= 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \int d^3 k_b d^3 k_c \delta_D(\mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_3 - \mathbf{k}_4) F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, \mathbf{k}_c] \delta_D(\mathbf{k}_b + \mathbf{k}_c) P_{k_b}^m + 2 p. \\
&= 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \int d^3 k_b F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, -\mathbf{k}_b] P_{k_b}^m + 2 p. \\
&= 3 \times (2\pi)^3 P_g^s(\mathbf{k}_1) F_s^{(1)}[\mathbf{k}_3] P_{k_3}^m \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \int d^3 k_b F_s^{(3)}[-\mathbf{k}_3, \mathbf{k}_b, -\mathbf{k}_b] P_{k_b}^m + 2 p.. \tag{B18}
\end{aligned}$$

From the last line, it is clear that this is also an unconnected term. The second term gives the tree level contribution

$$\begin{aligned}
& (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^{s(3b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\
&= \frac{6}{(2\pi)^6} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d^3 k_a d^3 k_b d^3 k_c \delta_D(\mathbf{k}_a + \mathbf{k}_b + \mathbf{k}_c - \mathbf{k}_4) F_s^{(3)}[\mathbf{k}_a, \mathbf{k}_b, \mathbf{k}_c] \\
&\quad \times (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_b) \delta_D(\mathbf{k}_3 + \mathbf{k}_c) P_{k_1}^m P_{k_2}^m P_{k_3}^m + 3 p. \\
&= 6 \times (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) F_s^{(3)}[-\mathbf{k}_1, -\mathbf{k}_2, -\mathbf{k}_3] P_{k_1}^m P_{k_2}^m P_{k_3}^m + 3 p.. \tag{B19}
\end{aligned}$$

Therefore, the full trispectrum at tree level is given by

$$T_g^s = T_g^{s(2b)} + T_g^{s(2c)} + T_g^{s(3b)}. \tag{B20}$$

**B4 Tetraspectrum definition**

The tetraspectrum is defined as

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5), \quad (\text{B21})$$

where the subscript ‘c’ stands for the connected part of the five points correlation function in Fourier space.

Drawing the diagrams as shown in Fry (1984), it is possible to see that the leading term of the connected part has order  $O(\delta_m^8)$ . There are two possible ways through which  $O(\delta_m^8)$  can be reached, either by having  $F_s^{(2)3}$  or  $F_s^{(2)}F_s^{(3)}$  terms:

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) &= \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \\ &\times \int d\mathbf{k}_c^3 d\mathbf{k}_d^3 \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] \int d\mathbf{k}_e^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] \langle \delta_1 \delta_2 \delta_a \delta_b \delta_c \delta_d \delta_e \delta_f \rangle + 59 \text{ p.} \\ &+ \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \\ &\times \int d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] \langle \delta_1 \delta_2 \delta_3 \delta_a \delta_b \delta_c \delta_d \delta_e \rangle + 19 \text{ p.} \end{aligned} \quad (\text{B22})$$

Starting from the first term that appears in the above expansion:

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^{s(a)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) &= \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) \\ &\times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] \langle \delta_1 \delta_2 \delta_a \delta_b \delta_c \delta_d \delta_e \delta_f \rangle + 59 \text{ p.} \\ &= 8 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_f^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_3) \delta_D(\mathbf{k}_c + \mathbf{k}_d - \mathbf{k}_4) \delta_D(\mathbf{k}_e + \mathbf{k}_f - \mathbf{k}_5) \\ &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_f) \delta_D(\mathbf{k}_2 + \mathbf{k}_a) \delta_D(\mathbf{k}_b + \mathbf{k}_c) \delta_D(\mathbf{k}_d + \mathbf{k}_e) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(2)}[\mathbf{k}_c, \mathbf{k}_d] F_s^{(2)}[\mathbf{k}_e, \mathbf{k}_f] P^m(k_1) P^m(k_2) P^m(k_b) P^m(k_c) + 359 \text{ p.} \\ &= 8 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \int d\mathbf{k}_b^3 d\mathbf{k}_d^3 \delta_D(-\mathbf{k}_2 + \mathbf{k}_b - \mathbf{k}_3) \delta_D(-\mathbf{k}_b + \mathbf{k}_d - \mathbf{k}_4) \delta_D(-\mathbf{k}_d - \mathbf{k}_1 - \mathbf{k}_5) \\ &\times F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_b] F_s^{(2)}[-\mathbf{k}_b, \mathbf{k}_d] F_s^{(2)}[-\mathbf{k}_d, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(k_b) P^m(k_b) + 359 \text{ p.} \\ &= (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) 8 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] \\ &\times F_s^{(2)}[-\mathbf{k}_2, \mathbf{k}_2 + \mathbf{k}_3] F_s^{(2)}[-\mathbf{k}_2 - \mathbf{k}_3, -\mathbf{k}_1 - \mathbf{k}_5] F_s^{(2)}[\mathbf{k}_1 + \mathbf{k}_5, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(|\mathbf{k}_2 + \mathbf{k}_3|) P^m(|\mathbf{k}_2 + \mathbf{k}_3|) + 359 \text{ p.} \end{aligned} \quad (\text{B23})$$

where the factor of 8 comes from permutations which do not influence the relations between the starting five wave vectors. On the contrary for each the original 60 permutations, there are six for which the relation between the initial wave vectors varies. Considering now the second term:

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^{s(b)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) &= \frac{1}{(2\pi)^9} F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) \\ &\times F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] \langle \delta_1 \delta_2 \delta_3 \delta_a \delta_b \delta_c \delta_d \delta_e \rangle + 19 \text{ p.} \\ &= 6 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 d\mathbf{k}_b^3 d\mathbf{k}_c^3 d\mathbf{k}_d^3 d\mathbf{k}_e^3 \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_4) \delta_D(\mathbf{k}_c + \mathbf{k}_d + \mathbf{k}_e - \mathbf{k}_5) \\ &\times \delta_D(\mathbf{k}_1 + \mathbf{k}_e) \delta_D(\mathbf{k}_2 + \mathbf{k}_d) \delta_D(\mathbf{k}_3 + \mathbf{k}_b) \delta_D(\mathbf{k}_a + \mathbf{k}_c) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] F_s^{(3)}[\mathbf{k}_c, \mathbf{k}_d, \mathbf{k}_e] P^m(k_1) P^m(k_2) P^m(k_3) P^m(k_a) + 119 \text{ p.} \\ &= 6 (2\pi)^3 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \int d\mathbf{k}_a^3 \delta_D(\mathbf{k}_a - \mathbf{k}_3 - \mathbf{k}_4) \delta_D(-\mathbf{k}_a - \mathbf{k}_2 - \mathbf{k}_1 - \mathbf{k}_5) \\ &\times F_s^{(2)}[\mathbf{k}_a, -\mathbf{k}_3] F_s^{(3)}[-\mathbf{k}_a, -\mathbf{k}_2, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(k_3) P^m(k_a) + 119 \text{ p.} \\ &= (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) 6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] \\ &\times F_s^{(2)}[\mathbf{k}_3 + \mathbf{k}_4, -\mathbf{k}_3] F_s^{(3)}[-\mathbf{k}_3 - \mathbf{k}_4, -\mathbf{k}_2, -\mathbf{k}_1] P^m(k_1) P^m(k_2) P^m(k_3) P^m(|\mathbf{k}_3 + \mathbf{k}_4|) + 119 \text{ p.} \end{aligned} \quad (\text{B24})$$

where the factor of 6 comes from permutations which do not influence the relations between the starting five wave vectors. On the contrary for each the original 19 permutations, there are six for which the relation between the initial wave vectors varies. From the above expressions of both  $T_{e,g}^{s(a)}$  and  $T_{e,g}^{s(b)}$ , it is possible to see that at leading order the tetraspectrum has order  $O(\delta_m^8)$ . In other words, it is proportional to the fourth power of the linear matter power spectrum.



### B5 Unconnected part of the five points correlation function

If one looks at the unconnected part of the five-point correlation function, the leading term has order  $O(\delta_m^6)$ . Below are derived the analytical expression for this unconnected part.

$$U_{5p,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \frac{1}{(2\pi)^3} \prod_{i=1}^4 F_s^{(1)}[\mathbf{k}_i] \int d\mathbf{k}_a d\mathbf{k}_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] \langle \delta_1 \delta_2 \delta_3 \delta_4 \delta_a \delta_b \rangle + 5 \text{ p..} \quad (\text{B25})$$

Using Wick's theorem, the six-points correlator can be approximated using in products of two points correlators. For example, for the first permutation one obtains

$$\begin{aligned} \langle \delta_1 \delta_2 \delta_3 \delta_4 \delta_a \delta_b \rangle &= 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_2 \delta_b \rangle \langle \delta_3 \delta_4 \rangle \leftrightarrow U_{5p}^1 \\ &+ 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_3 \delta_b \rangle \langle \delta_2 \delta_4 \rangle \leftrightarrow U_{5p}^2 \\ &+ 2 \times \langle \delta_1 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_2 \delta_3 \rangle \leftrightarrow U_{5p}^3 \\ &+ 2 \times \langle \delta_2 \delta_a \rangle \langle \delta_3 \delta_b \rangle \langle \delta_1 \delta_4 \rangle \leftrightarrow U_{5p}^4 \\ &+ 2 \times \langle \delta_2 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_1 \delta_3 \rangle \leftrightarrow U_{5p}^5 \\ &+ 2 \times \langle \delta_3 \delta_a \rangle \langle \delta_4 \delta_b \rangle \langle \delta_1 \delta_2 \rangle \leftrightarrow U_{5p}^6. \end{aligned} \quad (\text{B26})$$

Since all these terms are similar to each other, we compute below only the first one of them as an example on how to derive the others.

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) U_{5p}^1 &= \frac{2}{(2\pi)^3} \prod_{i=1}^4 F_s^{(1)}(\mathbf{k}_i) \int d\mathbf{k}_a d\mathbf{k}_b \delta_D(\mathbf{k}_a + \mathbf{k}_b - \mathbf{k}_5) F_s^{(2)}[\mathbf{k}_a, \mathbf{k}_b] (2\pi)^9 \delta_D(\mathbf{k}_1 + \mathbf{k}_a) \delta_D(\mathbf{k}_2 + \mathbf{k}_b) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) P^m(k_1) P^m(k_2) P^m(k_3) \\ &= 2 \times (2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_5) P^m(k_1) P^m(k_2) P^m(k_3) \\ &= 2 \times (2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_2] \delta_{125}^K \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P^m(k_1) P^m(k_2) P^m(k_3), \end{aligned} \quad (\text{B27})$$

where it is important to note that, in the last line, when substituting the Dirac's delta for the sum of three wave vectors with the sum of all five of them, a Kronecker's delta must be added to keep track of the original relation between the three  $\mathbf{k}$ 's. From the last line, one can immediately write another one of the six terms

$$\begin{aligned} (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) U_{5p}^5 &= 2(2\pi)^6 F_s^{(1)}[\mathbf{k}_1] F_s^{(1)}[\mathbf{k}_2] F_s^{(1)}[\mathbf{k}_3] F_s^{(1)}[\mathbf{k}_4] F_s^{(2)}[\mathbf{k}_2, \mathbf{k}_4] \delta_{245}^K \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P^m(k_2) P^m(k_4) P^m(k_3). \end{aligned} \quad (\text{B28})$$

Therefore, it is possible to write down the complete expression that takes into account all the permutations and terms as follows

$$U_{5p,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) = \sum_{i=1}^5 U_{5p,g}^s(\delta_i^{(2)}) \text{ where } U_{5p,g}^s(\delta_i^{(2)}) = U_{5p}^1 + U_{5p}^2 + U_{5p}^3 + U_{5p}^4 + U_{5p}^5 + U_{5p}^6. \quad (\text{B29})$$

The argument  $\delta_i^{(2)}$  identifies which wave vectors corresponds to the second-order perturbation expansion for each term.

## APPENDIX C: COVARIANCE TERMS DERIVATION

In this appendix, we present the derivation of the covariance terms, starting from the power spectrum one.

### C1 Covariance term: $C^{PP}$

As stated in the main text, the covariance matrix element can be computed as

$$C_g^{ss} \equiv \text{Cov} \left[ P_g^s(\bar{\mathbf{k}}_1), P_g^s(\bar{\mathbf{k}}_3) \right] = \left\langle \left( \hat{P}_{g,1}^s - \langle \hat{P}_{g,1}^s \rangle \right) \right\rangle \left\langle \left( \hat{P}_{g,3}^s - \langle \hat{P}_{g,3}^s \rangle \right) \right\rangle = \langle \hat{P}_{g,1}^s \hat{P}_{g,3}^s \rangle - \hat{P}_{g,1}^s \hat{P}_{g,3}^s. \quad (\text{C1})$$

The computation of the covariance matrix involves a four points correlator of  $\delta_g^s$  that can be expanded into its connected parts

$$\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle = \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle_c \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c + 2 \text{ perms.} + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c. \quad (\text{C2})$$

In this work, the following shorthand for the integrals will be also adopted

$$\int_{V_{k_i}^c} \frac{d\phi_i d\mathbf{k}_i^\perp d\mathbf{k}_i^\parallel}{2\pi \bar{k}_i^\perp \Delta k^2} \equiv \int_i. \quad (\text{C3})$$

*C1.1 C<sub>P</sub><sup>PP</sup> term*

Starting from the first term  $C_P^{PP}$  in the particular case in which the pairs are made of identical wave vectors (in this case, from the power spectrum estimator definition:  $\mathbf{k}_1 = \mathbf{k}_2$  and  $\mathbf{k}_3 = \mathbf{k}_4$ ), we have that

$$C_P^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) = \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle, \quad (C4)$$

it is straightforward to see that this terms cancels the second term of equation (C1). Considering now the other two possibilities for  $C_P^{PP}$

$$\begin{aligned} C_P^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \rangle \langle \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \rangle + 1p. \\ &= \frac{(2\pi)^6}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_2 + \mathbf{k}_4) P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) + 1p. \\ &= \frac{(2\pi)^6}{V_s^2 V_1^c V_2^c} (\delta_{13}^K \delta_{24}^K + \delta_{14}^K \delta_{23}^K) \int dV_1^c \int dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) \\ &= \frac{(2\pi)^3}{V_s V_1^c V_2^c} D_{1234}^P \int dV_1^c \int dV_2^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_2) \\ &= \frac{(2\pi)^3}{V_s V_1^c} D_{1234}^P \int_{\bar{k}_1^\perp - \frac{\Delta k}{2}}^{\bar{k}_1^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_1^\parallel - \frac{\Delta k}{2}}^{\bar{k}_1^\parallel + \frac{\Delta k}{2}} \frac{dk_1^\perp dk_1^\parallel}{\bar{k}_1^\perp \Delta k^2} P_g^s(\mathbf{k}_1)^2 \\ &\approx \frac{(2\pi)^2}{V_s \bar{k}_1^\perp \Delta k^2} D_{1234}^P P_g^s(\bar{\mathbf{k}}_1)^2 = \frac{2 \times (2\pi)^2}{V_s \bar{k}_1^\perp \Delta k^2} \delta_{13}^K P_g^s(\bar{\mathbf{k}}_1)^2, \end{aligned} \quad (C5)$$

where  $D_{1234}^P = \delta_{13}^K \delta_{24}^K + \delta_{14}^K \delta_{23}^K$  has been defined in order to take in account the additional permutation from the initial Dirac's deltas conditions, which now reduces to  $D_{1234}^P = 2 \times \delta_{13}^K$ . If, when obtaining an expression for the trispectrum, one considers also one-loop corrections, in addition to this Gaussian term on the diagonal of the power spectrum covariance matrix, one should add the unconnected terms encountered in Appendix B,  $T_g^{s(2a)}$  and  $T_g^{s(3a)}$ . This has been recently well described in Mohammed et al. (2017). As stated in Section 2.2, the error made with this tree level approximation for the diagonal term (C5) is around  $\sim 1$  per cent for  $k_{\max} = 0.2 \text{ Mpc}^{-1} h$  and up to  $\sim 9$  per cent for  $k_{\max} = 0.3 \text{ Mpc}^{-1} h$  (Taruya et al. 2008).

*C1.2 C<sub>T</sub><sup>PP</sup> term*

Proceeding in the same way, it is possible to compute as well the other term of the cross-correlation matrix, the one containing the trispectrum contribution.

$$\begin{aligned} C_T^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &= \frac{1}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \rangle \\ &= \frac{(2\pi)^3}{V_s^2 V_1^c V_3^c} \prod_{i=0}^4 \int dV_i^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4) T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\ &= \frac{(2\pi)^3}{V_s^2 V_1^c V_3^c} \delta_{34}^K \int dV_1^c dV_2^c dV_3^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2)^2 T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, -\mathbf{k}_3) \\ &= \frac{1}{V_s V_1^c V_3^c} \delta_{34}^K \int dV_1^c dV_2^c dV_3^c \delta_D(\mathbf{k}_1 + \mathbf{k}_2) T_g^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, -\mathbf{k}_3) \\ &= \frac{1}{V_s} \delta_{12}^K \delta_{34}^K \int \frac{dV_1^c}{V_1^c} \int \frac{dV_3^c}{V_3^c} T_g^s(\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_3, -\mathbf{k}_3) \\ &\approx \frac{1}{2\pi V_s} \int d\phi_{13} T_g^s(\bar{k}_1^\perp, \bar{k}_1^\parallel, \bar{k}_3^\perp, \bar{k}_3^\parallel, \phi_{13}). \end{aligned} \quad (C6)$$

Following what was done in Pielorz et al. (2010), i.e., substituting in here the previously derived expression for the trispectrum, this expression can be ‘simplified’ at tree level to

$$\begin{aligned} C_T^{PP}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3) &\approx \frac{1}{2\pi V_s} \int d\phi_{13} T_g^s(\bar{k}_1^\perp, \bar{k}_1^\parallel, \bar{k}_3^\perp, \bar{k}_3^\parallel, \phi_{13}) \\ &= \frac{1}{2\pi V_s} \int d\phi_{13} \left\{ 4 P_g^s(\mathbf{k}_1)^2 \left[ F_s^{(2)}[\mathbf{k}_1, -\mathbf{k}_+]^2 P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_-]^2 P^m(\mathbf{k}_-) \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + 4 P_g^s(\mathbf{k}_3)^2 \left[ F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_+]^2 P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_-]^2 P^m(\mathbf{k}_-) \right] \\
& + 8 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_3) \left[ F_s^{(2)}[\mathbf{k}_1, -\mathbf{k}_+] F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_+] P^m(\mathbf{k}_+) + F_s^{(2)}[\mathbf{k}_1, \mathbf{k}_-] F_s^{(2)}[\mathbf{k}_3, -\mathbf{k}_-] P^m(\mathbf{k}_-) \right] \\
& + 12 P_g^s(\mathbf{k}_1) P_g^s(\mathbf{k}_3) \left[ P_g^s(\mathbf{k}_1) F_s^{(3)}[\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_3] + P_g^s(\mathbf{k}_3) F_s^{(3)}[\mathbf{k}_1, \mathbf{k}_3, -\mathbf{k}_3] \right] \Bigg\}, \tag{C7}
\end{aligned}$$

where  $\mathbf{k}_+ = \mathbf{k}_1 + \mathbf{k}_3$  and  $\mathbf{k}_- = \mathbf{k}_3 - \mathbf{k}_1$  and  $P^m$  is the linear matter power spectrum and  $\phi_{13}$  is the difference between the two azimuthal angles of the two wave vectors. Therefore, this is the only term that requires an integration since this angular dependence defines the relation between the two wave vectors involved.

## C2 Covariance term: $\mathbf{C}^{\text{BB}}$

By definition, the covariance matrix element can be computed as

$$\begin{aligned}
\mathbf{C}^{\text{BS}} & \equiv \text{Cov} \left[ \mathbf{B}_g^s \left( \bar{k}_1^\perp, \bar{k}_2^\perp, \bar{k}_3^\perp, \bar{k}_1^\parallel, \bar{k}_2^\parallel \right), \mathbf{B}_g^s \left( \bar{k}_4^\perp, \bar{k}_5^\perp, \bar{k}_6^\perp, \bar{k}_4^\parallel, \bar{k}_5^\parallel \right) \right] \\
& = \left\langle \left( \hat{\mathbf{B}}_{g,123}^s - \langle \hat{\mathbf{B}}_{g,123}^s \rangle \right) \right\rangle \left\langle \left( \hat{\mathbf{B}}_{g,456}^s - \langle \hat{\mathbf{B}}_{g,456}^s \rangle \right) \right\rangle = \langle \hat{\mathbf{B}}_{g,123}^s \hat{\mathbf{B}}_{g,456}^s \rangle - \hat{\mathbf{B}}_{g,123}^s \hat{\mathbf{B}}_{g,456}^s. \tag{C8}
\end{aligned}$$

The six-points correlation function can be decomposed into all the possible combinations of connected parts

$$\begin{aligned}
\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle & = \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle_c \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c \langle \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 14 \text{ perms.} \\
& + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \rangle_c \langle \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 9 \text{ perms.} \\
& + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c \langle \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c + 14 \text{ perms.} \\
& + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \delta_g^s(\mathbf{k}_6) \rangle_c. \tag{C9}
\end{aligned}$$

Below we present the full derivation of the bispectrum Covariance matrix to leading order in Gaussian terms

$$\begin{aligned}
\mathbf{C}^{\text{BB}}(\bar{\mathbf{k}}_1, \bar{\mathbf{k}}_2, \bar{\mathbf{k}}_3; \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5, \bar{\mathbf{k}}_6) & = \frac{(2\pi)^4 \Delta k_3^\parallel \Delta k_6^\parallel}{V_s^2} \Lambda_{123}^{-1} \Lambda_{456}^{-1} \prod_{i=1}^6 \int_i \frac{dV_i^c}{V_i^c} F_{s,i}^{(1)} \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \delta_D(\mathbf{k}_4 + \mathbf{k}_5 + \mathbf{k}_6) \langle \delta_m(\mathbf{k}_1) \delta_m(\mathbf{k}_4) \rangle \langle \delta_m(\mathbf{k}_2) \delta_m(\mathbf{k}_5) \rangle \langle \delta_m(\mathbf{k}_3) \delta_m(\mathbf{k}_6) \rangle + 5 \text{ p.} \\
& = \frac{(2\pi)^{13} \Delta k_3^{\parallel 2}}{V_s^2 V_1^c V_2^c V_3^c} \Lambda_{123}^{-2} \delta_{14}^K \delta_{25}^K \delta_{36}^K \prod_{i=1}^3 \int_i F_{s,i}^{(1)2} P^m(k_i) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)^2 + 5 \text{ p.} \\
& = \frac{(2\pi)^{10} \Delta k_3^{\parallel 2}}{V_s V_1^c V_2^c V_3^c} \Lambda_{123}^{-2} D_{123456} \prod_{i=1}^3 \int_i P_g^s(\mathbf{k}_i) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \\
& = \frac{(2\pi)^{11} \Delta k_3^{\parallel 2}}{V_s V_1^c V_2^c V_3^c} \Lambda_{123}^{-1} D_{123456} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \int_{\bar{k}_i^\parallel - \frac{\Delta k}{2}}^{\bar{k}_i^\parallel + \frac{\Delta k}{2}} \frac{dk_i^\perp dk_i^\parallel}{V_i^c} P_g^s(\mathbf{k}_i) \delta_D\left(k_1^\parallel + k_2^\parallel + k_3^\parallel\right) \\
& = \frac{(2\pi)^5 \Delta k_3^\parallel}{V_s \bar{k}_1^\perp \bar{k}_2^\perp \bar{k}_3^\perp \Delta k_6^\parallel} \Lambda_{123}^{-1} D_{123456} \prod_{i=1}^3 \int_{\bar{k}_i^\perp - \frac{\Delta k}{2}}^{\bar{k}_i^\perp + \frac{\Delta k}{2}} \frac{dk_i^\perp}{\bar{k}_i^\perp \Delta k_i^\perp} \prod_{j=1}^2 \int_{\bar{k}_j^\parallel - \frac{\Delta k}{2}}^{\bar{k}_j^\parallel + \frac{\Delta k}{2}} \frac{dk_j^\parallel}{\Delta k_j^\parallel} P_g^s(\mathbf{k}_i), \tag{C10}
\end{aligned}$$

where  $D_{123456}$  has been used as a shorthand notation for all the possible permutations allowed by the fact that only pairs formed by wave vectors from different triplets survive, which are in total six. For all the other steps, we have used relations described previously. Finally, defining a bin average by using the thin shell approximation, it is possible to write down the final result

$$\mathbf{C}^{\text{BB}} = \frac{(2\pi)^5 \Delta k_3^\parallel}{V \bar{k}_1^\perp \bar{k}_2^\perp \bar{k}_3^\perp \Delta k_6^\parallel} \Lambda_{123}^{-1} D_{123456} \prod_{i=1}^3 F_s^{(1)2}[\bar{k}_i^\perp, \bar{k}_i^\parallel] P^m(\bar{k}_i^\perp, \bar{k}_i^\parallel). \tag{C11}$$

## C3 Cross-variance term: $\mathbf{C}^{\text{BP}}$

The cross-correlation part of the covariance matrix of the joint data vector  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$  can be computed as

$$\mathbf{C}^{\text{BP}} \equiv \text{Cov} \left[ \mathbf{P}_g^s(\bar{\mathbf{k}}_1), \mathbf{B}_g^s(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) \right] = \langle \hat{\mathbf{P}}_{g,1}^s \hat{\mathbf{B}}_{g,345}^s \rangle - \hat{\mathbf{P}}_{g,1}^s \hat{\mathbf{B}}_{g,345}^s. \tag{C12}$$

The computation of the covariance matrix involves a five points correlator of  $\delta_g^s$  that can be expanded in its connected parts

$$\begin{aligned}
\langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle &= \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \rangle_c \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c \\
&\quad + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \rangle_c \langle \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c + 5 \text{ p.} \\
&\quad + \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \rangle_c \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_5) \rangle_c + 2 \text{ p.} \\
&\quad + \langle \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_5) \rangle_c \\
&\quad + \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_{\text{unc.}} \\
&= \mathbf{C}_{m0}^{\text{BP}} + \mathbf{C}_{m1}^{\text{BP}} + \mathbf{C}_{m2}^{\text{BP}} + \mathbf{C}_{\text{Te}}^{\text{BP}} + \mathbf{C}_{\text{Up}}^{\text{BP}}.
\end{aligned} \tag{C13}$$

In the case of the five points correlation function, it has to be included also the unconnected part in the expansion. This is because while for an even number  $2n$ , the leading term of unconnected part of the  $2n$ -points correlation function is proportional to the product of  $n$  power spectra, for an odd number  $m$  this is no longer the case. As shown in last subsection of Appendix B, the unconnected part of the five points correlation function has a leading term proportional to the product of three power spectra. Therefore, in this case, the unconnected part needs to be considered since it has the same order of the other leading terms of the above expansion. From the last equation, it is possible to immediately that  $\mathbf{C}_{m0}^{\text{BP}}$  cancels out with the last term of equation (C12) and that  $\mathbf{C}_{m2}^{\text{BP}} = 0$  because the Dirac's deltas combination leads to terms having  $\delta_D(\mathbf{k}_5) = 0$ .

### C3.1 $\mathbf{C}_{m1}^{\text{BP}}$ term

Starting from the first term written previously, below we report the full derivation

$$\begin{aligned}
\mathbf{C}_{m1}^{\text{BP}}(\bar{\mathbf{k}}_1; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \frac{(2\pi)^2 \Delta k_5^\parallel}{V_s^2} \Lambda_{345}^{-1} \int dV_i^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \rangle_c \langle \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c + 5 \text{ p.} \\
&= \frac{(2\pi)^8 \Delta k_5^\parallel}{V_s^2} \Lambda_{345}^{-1} \int dV_i^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_2 + \mathbf{k}_4 + \mathbf{k}_5) P_g^s(\mathbf{k}_3) B_g^s(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
&= \frac{(2\pi)^5 \Delta k_5^\parallel}{V_s} \Lambda_{345}^{-1} \int dV_i^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P_g^s(\mathbf{k}_3) B_g^s(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
&= \frac{(2\pi)^5 \Delta k_5^\parallel}{V_s} \Lambda_{345}^{-1} \delta_{12}^K \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(-\mathbf{k}_2 + \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P_g^s(\mathbf{k}_3) B_g^s(\mathbf{k}_2, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
&= \frac{(2\pi)^5 \Delta k_5^\parallel}{V_s V_2^c} \Lambda_{345}^{-1} \delta_{12}^K \delta_{-23}^K \prod_{i=3}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) P_g^s(\mathbf{k}_3) B_g^s(\mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
&= \frac{(2\pi)^5 \Delta k_5^\parallel}{V_s V_2^c} \Lambda_{345}^{-1} \delta_{12}^K \delta_{-23}^K \prod_{i=3}^5 \int \frac{dk_i^\parallel dk_i^\perp}{2\pi \bar{k}_i^\perp \Delta k^2} 2\pi \Lambda_{345} \delta_D \left( k_3^\parallel + k_4^\parallel + k_5^\parallel \right) P_g^s(\mathbf{k}_3) B_g^s(\mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) + 5 \text{ p.} \\
&\approx \frac{(2\pi)^3}{V_s V_2^c} \delta_{-23}^K P_g^s(\bar{\mathbf{k}}_3) B_g^s(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) + 5 \text{ p..}
\end{aligned} \tag{C14}$$

Note that only three of the six permutations are different terms, since the other three are just obtained by switching  $\mathbf{k}_1$  with  $\mathbf{k}_2$  which does not change the final result. Therefore, the final expression for this term can be written as

$$\mathbf{C}_{m1}^{\text{BP}} = 2 \times \frac{(2\pi)^3}{V_s V_2^c} (\delta_{-13}^K + \delta_{-14}^K + \delta_{-15}^K) P_g^s(\bar{\mathbf{k}}_1) B_g^s(\bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5). \tag{C15}$$

Note that the argument of the power spectrum depend on the wave vector selected by the Kronecker Delta.

### C3.2 $\mathbf{C}_{\text{Te}}^{\text{BP}}$ term

The other contribution at tree level to the connected five points correlator resulting in the cross-covariance term between power spectrum and bispectrum is the one proportional to the tetraspectrum defined before, we get

$$\begin{aligned}
\mathbf{C}_{\text{Up}}^{\text{BP}}(\bar{\mathbf{k}}_2; \bar{\mathbf{k}}_3, \bar{\mathbf{k}}_4, \bar{\mathbf{k}}_5) &= \frac{(2\pi)^2 \Delta k_5^\parallel}{V_s^2} \Lambda_{345}^{-1} \int dV_i^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_3) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \langle \delta_g^s(\mathbf{k}_1) \delta_g^s(\mathbf{k}_3) \delta_g^s(\mathbf{k}_2) \delta_g^s(\mathbf{k}_4) \delta_g^s(\mathbf{k}_5) \rangle_c \\
&= \frac{(2\pi)^5 \Delta k_5^\parallel}{V_s^2} \Lambda_{345}^{-1} \int dV_i^c \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_1 + \mathbf{k}_2) \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5)
\end{aligned}$$



$$\begin{aligned}
&= \frac{(2\pi)^5 \Delta k_5^{\parallel}}{V_s^2} \Lambda_{345}^{-1} \delta_{12}^K \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5)^2 T_{e,g}^s(\mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5) \\
&= \frac{(2\pi)^2 \Delta k_5^{\parallel}}{V_s} \Lambda_{345}^{-1} \prod_{i=2}^5 \int \frac{dV_i^c}{V_i^c} \delta_D(\mathbf{k}_3 + \mathbf{k}_4 + \mathbf{k}_5) T_{e,g}^s(\mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4, \mathbf{k}_5).
\end{aligned} \tag{C16}$$

Since the leading order of the tetraspectrum is higher than the one considered in this work, the above covariance term has not been included in the numerical computation. For what concerns, the unconnected part of the five-point correlation function described in the end of Appendix B, it represents loop correction to the standard power spectrum times bispectrum cross-covariance  $C_{m1}^{\text{BP}}$ . Coherently with what done in the same case for the trispectrum, it has also not been included in the numerical computations.

#### APPENDIX D: COMPRESSED COVARIANCE MATRIX

We run the MCMC not on the original data vector, e.g.  $\mathbf{B}_g^s$  or  $[\mathbf{P}_g^s + \mathbf{B}_g^s]$  but on their compressed version obtained through K-L weighting. Given a set of parameters  $\{\theta_1, \dots, \theta_m\}$ , a data vector  $\mathbf{x} = \{x_1, \dots, x_n\}$  with mean  $\langle \mathbf{x} \rangle$  the set of weights with respect to each parameter  $\theta_i$  is given by

$$\mathbf{b}_i = \text{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}, \tag{D1}$$

where  $\text{Cov}(\mathbf{x})$  is the covariance matrix relative to the data vector  $\mathbf{x}$ . Assuming the covariance matrix to be approximately independent of the cosmology, then for each considered parameter, the original data vector is compressed to a single scalar

$$y_{\theta_i} = \mathbf{b}_i \cdot \mathbf{x}. \tag{D2}$$

Therefore stacking all the weighting vectors as rows of an  $m \times n$  matrix  $\mathbf{B}$ , the new  $m$ -dimensional data vector will be

$$\mathbf{y} = \mathbf{B}\mathbf{x}. \tag{D3}$$

For what concerns, the covariance matrix for the new data vector, it will be related to the original one by

$$\begin{aligned}
\text{Cov}_{y,i,j} &= \text{Cov}[y_i, y_j] = \text{Cov} \left[ \sum_k^n b_i^k x_k, \sum_j^n b_j^l x_l \right] \\
&= \sum_k^n \sum_j^n b_i^k b_j^l \text{Cov}[x_k, x_l] \\
&= \mathbf{b}_i^T \cdot \text{Cov}_x \cdot \mathbf{b}_j,
\end{aligned} \tag{D4}$$

Therefore when running the MCMC using  $\mathbf{y}$  as data vector, the natural logarithm of likelihood will be proportional to

$$\log \mathcal{L} \propto -\frac{1}{2} (\mathbf{y} - \mathbf{y}_{\text{fid.}})^T \text{Cov}_y^{-1} (\mathbf{y} - \mathbf{y}_{\text{fid.}}) \tag{D5}$$

where  $\mathbf{y}_{\text{fid.}}$  is the compressed data vector obtained using the fiducial values of the cosmological parameters. In our case of interest, this will be applied to  $\mathbf{B}_g^s$  and  $[\mathbf{P}_g^s, \mathbf{B}_g^s]$ . An additional option consists in leaving  $\mathbf{P}_g^s$  uncompressed, compressing only the bispectrum. In that case, the cross-variance term would be given by

$$\text{Cov}_{ij}^{P y_B} = \text{Cov}[\mathbf{P}_g^s(\mathbf{k}_i), y_j] = \text{Cov}[\mathbf{P}_g^s(\mathbf{k}_i), \mathbf{B}_g^s] \cdot \mathbf{b}_j. \tag{D6}$$

Putting everything together the full covariance matrix is

$$\text{Cov} = \begin{bmatrix} \text{Cov}^{PP} & \text{Cov}^{P y_B} \\ \text{Cov}^{P y_B} & \text{Cov}^{y_B y_B} \end{bmatrix}, \tag{D7}$$

where  $\text{Cov}^{y_B y_B}$  is the covariance matrix for the compressed data vector derived in eq. (D4) in the case of  $\mathbf{x}$  being the bispectrum.

#### APPENDIX E: WEIGHTS ORTHOGONALIZATION

As stated in Zablocki & Dodelson (2016), each compressed data set  $y_i$  contains all the information regarding the parameter  $\theta_i$ , but at the same time, it will have some sensitivity to the other parameters. It is possible to remove this sensitivity by marginalizing at linear level over these other parameters. This is done by orthogonalizing the weighting vectors for all the individual parameters. If the model is parametrized by  $m$  parameters  $\theta_i$ , compressing with respect to all these will return  $m$  different linear combinations  $y_i$  (or in other words  $m$  weighting vectors). Therefore, it is possible to define another set of new scalars  $y'_i$  given by a linear combination of the original ones, for example:

$$y'_1 = c_1 y_1 + c_2 y_2 + \dots + c_m y_m, \tag{E1}$$

with the constants  $c_i$  such that  $y'_1$  contains all the information regarding  $\theta_1$ , while at the same time removing all the sensitivity with respect to the other parameters at linear level. This means that when taking the derivative with respect to the second parameter  $\theta_2$  we are looking for  $c$  coefficients such that:

$$\frac{dy'_1}{d\theta_2} = c_1 \frac{dy_1}{d\theta_2} + c_2 \frac{dy_2}{d\theta_2} + \dots + c_m \frac{dy_m}{d\theta_2} = c_1 \sum_j b_1^j \frac{dx_j}{d\theta_2} + c_2 \sum_j b_2^j \frac{dx_j}{d\theta_2} + \dots + c_m \sum_j b_m^j \frac{dx_j}{d\theta_2} = c_1 F_{12} + c_2 F_{22} + \dots + c_m F_{m2} = 0, \quad (\text{E2})$$

where for the Fisher information matrix elements, it has been used equation (21) together with the expression for the weights  $\mathbf{b}_i = \mathbf{Cov}^{-1} \langle \mathbf{x} \rangle_{,i}$ . Imposing this for the derivative with respect to all the  $m - 1$  parameters beside  $\theta_1$ , it gives the following matrix problem to be solved in order to get the  $m - 1$  coefficients needed to compute  $y'_1$ :

$$\begin{bmatrix} F_{22} & F_{23} & \dots & F_{2m} \\ F_{32} & F_{33} & \dots & F_{3m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ F_{m2} & F_{m3} & \dots & F_{mm} \end{bmatrix} \begin{bmatrix} c_2 \\ c_3 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{bmatrix} = \begin{bmatrix} -F_{12} \\ -F_{13} \\ \cdot \\ \cdot \\ \cdot \\ -F_{1m} \end{bmatrix}. \quad (\text{E3})$$

Solving this equation returns  $m - 1$  unique constants for the  $c_i$  with  $i > 1$ .  $c_1$  can be set equal to one. Once the  $c$ 's are known,  $y'_1$  can be obtained using equation (E1), which corresponds to defining the new weighting vector for the parameter  $\theta_1$ :

$$\mathbf{b}'_1 = \mathbf{b}_1 + c_2 \mathbf{b}_2 + \dots + c_m \mathbf{b}_m, \quad (\text{E4})$$

where  $\mathbf{b}'_1$  is the orthogonalized weighting vector for  $\theta_1$ . The same procedure can be repeated for all the other parameters and relative weighting vectors. In general, for the mode with  $i = \alpha$ , the coefficients are determined by the general equation:

$$\sum_j F'_{\alpha,ij} c_j = -F_{\alpha i}, \quad (\text{E5})$$

where  $F'_\alpha$  is the Fisher matrix with row and column  $\alpha$  removed.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.